

Mad cows cast long shadows

Scientists may escape the worst of the flak from Britain's BSE inquiry, but they ignore its lessons at their peril.

After 34 long months, the public inquiry into Britain's BSE epidemic is to be released this week. Its 16 volumes will make uncomfortable reading. The report is expected to describe how attempts to protect the interests of the agriculture industry were allowed to override concerns about public health — and how, ultimately, both came to suffer. It seems likely that government ministers from the late 1980s to the mid-1990s, plus a number of senior civil servants then at the Ministry of Agriculture, Fisheries and Food (MAFF), will be sharply criticized for their roles in shaping this course of events.

The inquiry is also expected to address the adequacy of the system of providing, and acting on, scientific advice to government. It will examine how a lack of scientific knowledge about the risks posed to human health by an emerging disease of cattle became transformed in ministers' statements into assurances that there was no cause for concern. It would be harsh if the scientists who advised the politicians are themselves censured. But even if the inquiry judges them to have been blameless, there is much for the scientific community to learn from the BSE affair.

From evidence given to the inquiry, and the experience of *Nature's* staff in reporting on the epidemic, a sorry tale emerges. It is a story of MAFF concentrating government funding into a few 'trusted' laboratories — not necessarily those with the most appropriate expertise (see page 932). The ministry also held a veto over the publication of results from this research. And this journal has previously criticized a culture of secrecy within MAFF that prevented independent groups from analysing detailed epidemiological data (see *Nature* 383, 463; 1996).

MAFF also promulgated the view that BSE was caused by the same agent as a related disease of sheep called scrapie, even though the evidence was equivocal. This was a comforting assumption, as scrapie had been present in British sheep for more than 200 years, with no evidence of any risk to human health. When evidence began to emerge that BSE

was distinct from any known strain of scrapie, the official line did not change significantly. Today, experts say that it is impossible to determine whether the BSE agent crossed over from sheep, arose *de novo* in cows or had been present subclinically in cattle for many years. But it seems clear that it spread through British herds through the practice of feeding cows on meat and bone-meal derived from cattle carcasses.

In their defence, ministers of the day have pointed out that the precautionary measures taken in the late 1980s went further than the scientific advice they were given. But having taken the view that BSE was likely to pose no greater risk than scrapie, it is clear that these measures were not strictly enforced. Epidemiological data reveal that infected tissues continued to find their way into cattle-feed, even though the practice of feeding cows to cows had been banned since July 1988. The extent to which traces of potentially infective bovine central nervous tissue continued to enter the human food chain is unknown.

It is easy, with the benefit of hindsight, to criticize scientists for not making more of a fuss about the control that MAFF exerted over research, and for failing to dispel the continuing 'BSE is scrapie' complacency. But to do so would ignore the political and financial climate of the day. In the late 1980s and early 1990s, British agricultural and veterinary research was being 'restructured' — a euphemism for being cut to the bone. No surprise, then, that many scientists were keeping their heads down, not wishing to be identified as troublemakers.

Given the damage caused by the BSE affair, however, we now know the dangers of keeping silent. Of course, the main victims are those whose lives have been tragically cut short by a horrific disease. But science, too, has suffered. With ministers having consistently claimed that they were following the best scientific advice, even while subtly misrepresenting its message, scientists — particularly those working for government — have come to be seen by the British public as part of the problem. It will take much work to regain public trust. ■

Recognition for mathematics is overdue

The prospect of a boost in funding for US mathematicians should be warmly welcomed by the entire scientific community.

In proposing a major new initiative in mathematics (see page 931), Rita Colwell, director of the National Science Foundation (NSF), is recognizing problems that have been apparent to US mathematicians for some time. If enacted, its positive ramifications will extend beyond university mathematics departments.

The United States enjoys global leadership in many branches of mathematics, but this conceals some festering problems. First are weaknesses in its educational system. According to the Third International Mathematics and Science Study, US students do quite well in mathematics at age 9, rather badly at 13 and abysmally at 17. US universities also do poorly at recruiting American students into mathematics, especially at the graduate level. Twenty years ago, three-quarters of US mathematics PhDs were Americans: today, more than half are foreign-born.

For those who persist into academia, grant support is paltry by US standards. And agencies that once gave generous support for university mathematics, such as the departments of energy and

defence, now leave the NSF to foot two-thirds of the bill.

Yet mathematicians are much in demand. Their research has been applied in search engines and other Internet tools, and underpins progress in other disciplines. Thanks to the rise of genomics and structural biology, and the realization that modelling can yield valuable insights into biological systems, this is now as true in the life sciences as in the physical ones.

Colwell is aware of these challenges. In an ideal world, she says, the NSF's \$125 million mathematics budget would double in two years, and double again thereafter. Progress towards these goals would allow bigger grants, better linkages with other disciplines, new research institutes and improved ties between universities and high schools.

If this wins the backing of Congress and the next US administration, the whole of science — and society at large — stands to benefit. The initiative would concentrate on aspects of mathematics — such as the study of dynamic systems and the modelling of uncertainty — that will help tackle problems that affect us all. ■





Danger signs
New data show
growing spread of
coral bleaching
p932



Icy reception
Astronomers protest
over delay to Pluto
mission
p933



Backing biotech
Germany endorses
Europe's plans for
gene patents
p934



Population study
Singapore launches
nationwide disease
database
p935

NSF calls for funding boost in a bid to reverse decline in maths

Paul Smaglik, Washington

The US National Science Foundation (NSF) needs to more than triple its commitment to mathematics over the coming years to reverse the subject's decline and meet the growing needs of other disciplines, Rita Colwell, the director of the agency, told the National Science Board last week.

Boosting the agency's maths budget from \$106 million in 2000 to between \$400 million and \$500 million by 2007 would fund a campaign to make mathematics a more attractive profession, she said.

The proposal is expected to be a central part of the agency's next funding bid to Congress. It comes at a time when fewer graduate students are studying maths and the supply of foreign mathematicians is drying up.

The resulting shortage threatens the progress of a number of NSF initiatives, such as those in nanotechnology, biocomplexity and information technology. A lack of mathematicians could also delay analysis of the genomic and proteomic data emerging from

sequencing and related programmes.

Mathematics "underpins" all these efforts, Colwell told the National Science Board, which is formally responsible for the NSF's activities. But, she added, the discipline is being undermined by a lack of funding.

The average mathematics grant is now \$30,000 a year for three years. Colwell wants to extend this to five years and to increase the award's size — although she declined to give a figure. Without extra support, she warns, prospective mathematicians will move to other disciplines. The initiative would also provide more money for interdisciplinary research, and for training and education.

Robert Richardson, vice-provost for research and professor of physics at Cornell University, Ithaca, New York, and a member of the science board, agreed. He said that the "miserable grant size" of NSF mathematics awards deters people from following the subject at graduate level.

According to Philippe Tondeur, head of the NSF mathematics directorate and a



The way ahead: teaching needs more support.

maths professor at the University of Illinois at Urbana-Champaign, many undergraduates prefer to move into lucrative jobs in computer programming. Low stipends and inadequate grant support put them off a graduate education in mathematics, he says.

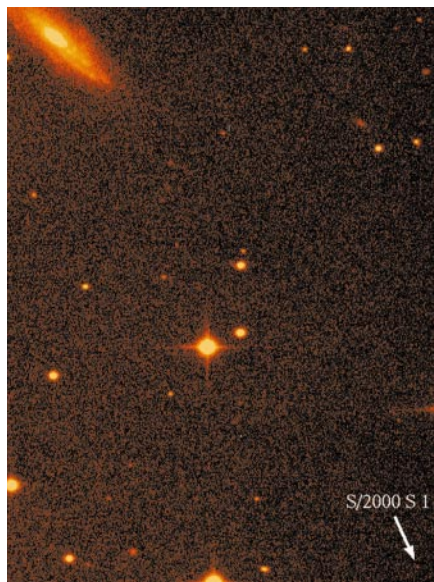
The proportion of US citizens among full-time maths graduate students in the United States dropped by 26.5% between 1992 and 1999 at the nation's top universities, according to figures from the American Mathematical Society.

Until recently, the United States had compensated for such shortages by bringing in mathematicians from China and Russia, just as previous shortages had been relieved by recruiting from Europe. But Tondeur says that many Chinese mathematicians are returning to their native country, and the Russian pool of talent is drying up — especially as the country's maths and science infrastructure crumbles.

Samuel Rankin, associate executive director of the American Mathematical Society, says the extra money proposed by the NSF will be good news — if it happens. "The mathematical community has to use Dr Colwell's statement and sentiment to help leverage this kind of increase for mathematics," he says.

Colwell also wants to raise the NSF stipends for graduate students from \$16,800 to \$25,000. To accomplish this while accommodating the NSF's many new and proposed initiatives, the agency will almost certainly need to double its budget over five years.

Four more moons for Saturn



David Adam

Saturn's extended family has grown larger. Astronomers have found four more moons orbiting the planet, taking its total to 22.

The moons are irregular and in orbits about 15 million kilometres from Saturn — much further than its regular moons (see left). Their discovery puts Saturn back in the lead as the planet with the most satellites.

The moons were spotted during the past two months by several telescopes. They were due to be announced at this week's meeting of the American Astronomical Society's Division for Planetary Science in Pasadena, California.

Brett Gladman, an astronomer at the Observatoire de la Côte d'Azur, France, first noticed two of the objects in images from the European Southern Observatory's 2.2-metre telescope in Chile. The others were found using a larger instrument in Hawaii.

► <http://pinks.physics.mcmaster.ca/Saturn>

UK ministry under fire over handling of BSE research

David Dickson, London

Britain's Ministry of Agriculture, Fisheries and Food (MAFF) was accused this week of hindering research into bovine spongiform encephalopathy (BSE) — mad cow disease — during the 1980s.

Maitland Mackie, former chairman of the animals committee of the then Agricultural and Food Research Council, said that MAFF awarded BSE-research grants to scientific “novices” working in its own laboratories, rather than to those in independent institutions with long experience in related areas.

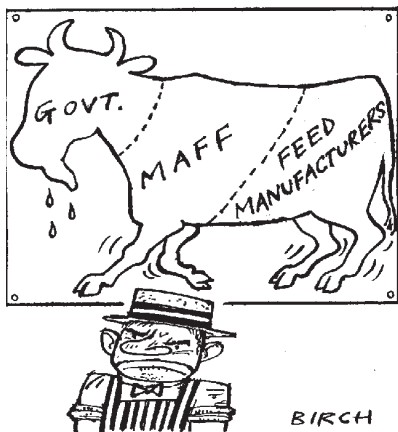
Mackie's comments came shortly before the publication this week of the conclusions of a public inquiry led by Lord Phillips into the BSE outbreak.

The inquiry is expected to distribute blame widely among MAFF officials and their political bosses for giving excessive weight to the interests of the food industry in their handling of information about the spread of BSE and its likely implications for human health.

The Phillips report is also expected to look critically at the impact of cuts to agricultural research in the 1980s, and the efforts by MAFF to exert control over the results that were published.

In evidence to the inquiry, for example, Alan Dickinson, head of the Neuropathogenesis Unit in Edinburgh between 1981 and 1987, had claimed that the overall research funds “were given disproportionately to those without previous experience”, with the result that his own “chronically underfunded” unit had had its work “seriously hindered”.

Speaking this week in a radio interview with the BBC, Mackie argued that MAFF's decision to channel funding to its own institutes had held up vital research into BSE by two years. ■



Global warming identified as main threat to coral reefs

Peter Pockley, Bali

A report released this week paints a bleak picture of the condition of the world's coral reefs. And it names global warming as the culprit behind much of the damage.

Only two years after a survey of the world's coral reefs found 11% had been destroyed before 1998, a more extensive assessment by 80 countries of their own reefs — most for the first time — has raised the total to 27% “effectively lost” by late 2000.

The latest measurements come from the Global Coral Reef Monitoring Network, the operational unit of the International Coral Reef Initiative, and they take account of the massive bleaching of corals that took place in 1998. The network identified a further 14% of reefs that are at “critical” risk of being lost within 2–10 years, and 18% more as being “threatened” in 10–30 years.

The details were released this week at the ninth International Coral Reef Symposium in Bali, Indonesia, an island that sits at the centre of the archipelago harbouring the greatest diversity of coral species.

The coordinator of both surveys, Clive Wilkinson of the Australian Institute of Marine Science (AIMS), says the agenda up to 1998 had primarily been to curb human impacts on reefs. Damage is now documented from over-fishing for subsistence and export, from the live fish trade, from fishing by explosion and poisoning, as well as from sewage, and excessive sediments and nutrients running off urban and agricultural development.

Although these influences remain, the new report argues that the main problem has now become global climate change, to which corals are proving acutely sensitive. Terry Done, also of AIMS and president of the con-



Damage done: bleached coral in Indonesia.

ference, joined Wilkinson in stressing the need to curb greenhouse gas emissions.

The problems are most severe in the wider Indian Ocean, where there has been 59% loss but “reasonable chances of recovery for remote reefs”, according to the report. Next in order are reefs in the Middle East (35% damage and “low chances for short-term recovery”), and Southeast and East Asia (34% loss “with reasonable chances for slow recovery on remote reefs”).

The Caribbean/Atlantic reported a loss of 22%, mostly from human impact — but with relatively fast recovery. In contrast, the extensive reefs in the Pacific and off Australia are reported as being “in reasonably good health with a positive outlook unless global climate change events like those of 1998 strike these areas”.

Wilkinson says it is significant that countries are “no longer in a state of denial and are appealing for help”, as exemplified by Indonesian officials at the conference. ■

♦ <http://coral.aoml.noaa.gov/gcrmn/gcrmn.html>

Carbon sequestration gains support

Colin Macilwain, Washington

Princeton University has been awarded \$15 million from the petrochemical company BP and \$5 million from the car maker Ford to set up a project to study carbon sequestration and other possible solutions to the problem of global climate change.

The project will try to find measures that will have limited environmental impact and which avoid “prohibitive economic costs” and “prohibitive disruptions in patterns of energy consumption”. These are expected to include work on the value of carbon sequestration in underground geological formations and the

possibility of switching to hydrogen fuels.

The research funding will be spread over ten years, under a Carbon Mitigation Initiative. The university describes it as the largest corporate grant in its history.

The initiative was due to be announced yesterday (25 October) in New York by BP chairman Sir John Browne and Harold Shapiro, president of Princeton.

Carbon sequestration has become a popular subject recently, despite doubts about its feasibility, as it has become apparent that the United States is not likely to act to reduce its carbon emissions. ■

US Geological Survey needs more resources, says draft report

Rex Dalton, San Diego

Research programmes at the US Geological Survey (USGS) should be buttressed with added funding, new staff and independent advisory committees, according to the draft of a report by the National Academy of Sciences (NAS).

The recommendations, a final version of which could be released as soon as this week, are intended to strengthen an agency that is widely seen as the US government's premier authority on Earth sciences.

The report warns that, unless personnel and financial issues are dealt with soon, "the USGS may be unable to meet all of its mission goals, respond rapidly to new challenges and opportunities, and [become] the nation's Earth-resource science and information management agency".

Already more than a year overdue, the report is the work of a panel of experts brought together by the academy's National Research Council to examine a 120-year-old agency that has been buffeted by political and economic forces in recent years.

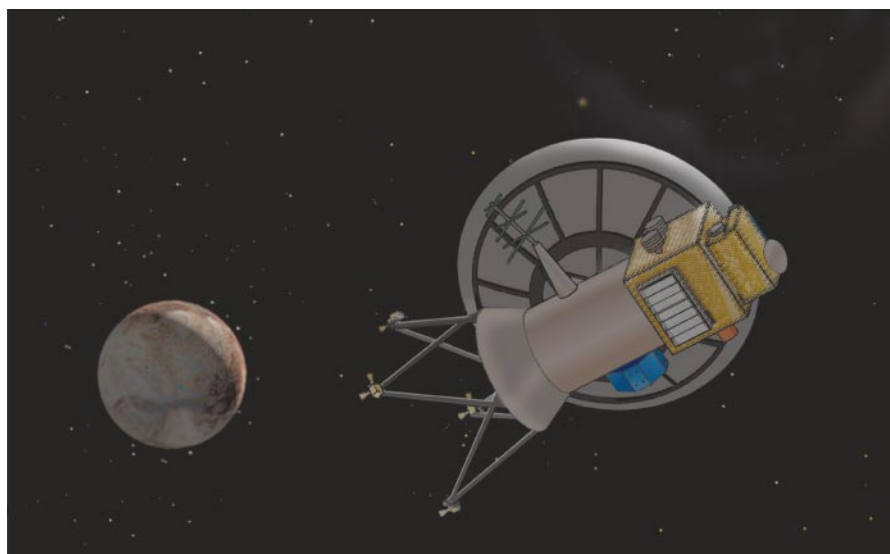
In the mid-1990s, Republicans in Congress tried to close the agency, which has about 10,000 employees and an annual budget of about \$1.1 billion, and is concerned with geology, water resources, mapping and biological resources.

The Republican bid grew out of conflict with President Bill Clinton's administration over environmental policy. As a compromise, the agency's staff was cut and a contract reached with the NAS to help chart a course for the future.

The draft report points out that, although the agency has recently built "a solid foundation on which to plan its future", a number of improvements are still needed to meet its long-term goals.

To ensure rigorous review of research projects, for example, it recommends that the agency "establish and make extensive use of external advisory committees" at national and regional levels. Such a move would be welcomed both inside and outside the USGS, as it would be seen as improving the scientific quality of its research.

The report also calls for greater emphasis on "multiscale, multidisciplinary projects" that integrate Earth and environmental problems. And it says that, to become more of "an information management" agency, the USGS "should shift from a more passive role of study and analysis to one that seeks actively to convey information in ways that are responsive to the social, political and economic needs". ■



Not so fast: on its new schedule, a revised Pluto-Kuiper Express will not reach Pluto until 2020.

Fan mail demands reprieve for delayed US Pluto mission

William Triplett, Washington

If you think Americans don't pay much attention to the least-explored planet in the Solar System, think again. Supporters of the US space agency NASA's Pluto-Kuiper Express mission — postponed last month for financial reasons — are letting their congressmen know about their unhappiness.

Last week, Louis Friedman, executive director of the Planetary Society in Pasadena, California, an organization of space-exploration enthusiasts, and television personality Bill Nye, a member of its board of directors, visited Washington to deliver 10,000 postcards and letters that the group has received from members who want the mission reinstated without delay.

Nye went to the office of Dana Rohrabacher (Republican, California), chairman of the space and aeronautics subcommittee of the House of Representatives science committee, to deliver 5,000 of the missives. Friedman took the other 5,000 to the office of Bill Frist (Republican, Tennessee), chair of the Senate science, technology and space subcommittee.

The key issue is timing. The Pluto-Kuiper Express, a space probe intended to take readings of Pluto's fragile atmosphere, was originally scheduled to be launched in 2004 and to arrive in 2012. On the new schedule, announced in September, the spacecraft will not reach Pluto until 2020.

Some scientists say this will be too late to provide the data they are hoping to collect. The further the planet travels from the Sun, the colder it gets — by 2020, Pluto's atmosphere may have frozen, and it would be about 200 years before it thaws again.

Another problem with the delay is that, at

the original launch date, the planet Jupiter would have given a valuable gravity assist, giving the spacecraft a slingshot-like push as it flew by. There will be no such assist under the new schedule.

NASA's decision to postpone the launch — taken shortly after the agency announced it was planning more extensive, and expensive, surface exploration of Mars — immediately freed up the \$500 million that the Mars project was expected to cost.

But Nye is said to have told Rohrabacher that NASA had not been totally open about the reasons for the delay. According to Susan Lendroth, events and communications manager for the Planetary Society, Rohrabacher told Nye that he "would look into it".

Ricardo Bernal, a member of Rohrabacher's staff, says the congressman plans to talk to NASA officials in January, when the new Congress convenes. "He wants to look into it very carefully to find out exactly why the mission was put on hold," says Bernal.

If the reasons turn out to be purely financial, Rohrabacher will want to know "exactly how much more will be needed", says Bernal. "He'll also want to know if there might be other options, such as a smaller probe."

Ed Weiler, NASA's associate administrator for space science, says the decision was "absolutely based on budgetary concerns". Referring to Nye's meeting with Rohrabacher, Weiler says he thinks the Planetary Society "did a great job of showing support for the Pluto mission, but they were literally a day late and a dollar short". NASA's appropriations bill for the fiscal year 2001 was passed by the house an hour after the meeting. ■

► <http://planetarium.org>

Germany gives green light to gene patents

Quirin Schiermeier, Munich

The German cabinet last week endorsed the patenting of human genes, when it approved a European Union (EU) directive on the legal protection of biotechnological inventions.

The move is a significant step towards the directive's adoption across Europe. It was approved at the European level in 1998 after almost ten years of preparation, and is intended to harmonize biotechnology patents in the EU (see *Nature* 388, 314; 1997). But ethicists and scientists still object to some clauses — albeit on different grounds.

Controversy surrounds a clause stating that patents can be granted on an "element isolated from the human body or otherwise produced by means of a technical process ... even if the structure of that element is identical to that of a natural element". The clause says that such elements can include "the sequence or partial sequence of a gene".

Some say this violates the human right to freedom and protection of the body from ownership. The Netherlands has filed a complaint against the directive at the European Court of Justice on these grounds.

"In its current form [the directive] would lead inexorably to the commercialization of genetic knowledge," says Jean-François Mattei, a Marseilles-based geneticist and member of the French parliament. Mattei will submit a petition calling for a moratorium to President Jacques Chirac next month.

Although EU members are legally obliged to adopt the directive and, where necessary, adapt their legislation to fit it, France and several other member states have yet to



Patent protest: Greenpeace have been orchestrating German objections to the patent law revisions.

do so. Those that have include Denmark, the Republic of Ireland, Finland and Britain.

Discussion is particularly heated in France, because the directive goes against its national patent laws. These state that "the human body, its parts and products, and knowledge of entire or partial structure of the gene as such" cannot be patented.

Many scientists' concerns centre on the breadth of gene patents. Some have warned that the development of drugs and therapies could be blocked, or prices increased, if gene patents could cover all possible functions of a sequence (see *Nature* 406, 111; 2000).

The ministerial comments accompanying the German bill, which are intended to guide its implementation, address this concern. They say that the German and European patent offices should "restrict a patent to those parts of a gene substantial for the function described in the application".

Applicants would need to give a concrete description of the function of a gene, and protection would generally be awarded on smaller sequences. This would minimize cases where additional patentable functions are assigned to already protected gene sequences.

But Axel Kahn, director of the INSERM Laboratory of Research on Genetics and Molecular Pathology at the Cochin Institute of Molecular Genetics in Paris, doubts that national comments will solve the problem.

"I regret that the directive contains no unambiguous clause saying that the use of DNA sequences for everything that was not initially described remains free," he says. Kahn points out that there is no reason for the European Patent Office to follow the guidelines in the German law.

But Joseph Straus, head of the department for European patent law at the Max Planck Institute for Foreign and International Patent, Copyright and Competition Law in Munich, is confident that European patent offices will respect the research community's desire for narrow gene patents.

The biotechnology industry in Germany and abroad has welcomed the government's decision. "Those European governments that have agreed on the directive are now reluctant to adopt it," says Bo Hammer Jensen, who chairs the intellectual property rights panel at EuropaBio, the association of European bio-industries. "We are happy that Germany has finally decided to take it on."

Human Frontier repatriation boost

Researchers awarded postdoctoral fellowships under the Human Frontier Science Program (HFSP) to study abroad are to get financial incentives to return home when the grant expires. The move reflects concern that a significant proportion are staying in the country that they have visited — particularly the United States.

HFSP fellowships are to be extended from two to three years, and grant-holders will be able to spend the third year back home. The HFSP has also proposed the creation of a new career development award. This would allow repatriated fellows to apply for a two-year extension to their fellowship, including US\$50,000 to build up an independent research group at home.

The HFSP was set up in 1989 to improve international collaboration in molecular biology and neuroscience. It now has a



Wiesel: money will improve training.

budget of about \$50 million per year. But 70% of HFSP fellows who move to the United States remain there after completing their fellowship (see *Nature* 399, 398; 1999).

Under the proposed new scheme, the number of fellowships awarded each year would fall by one-third, to 110.

But Torsten Wiesel, secretary-general of the HFSP and president emeritus of the Rockefeller University in New York, hopes that the new rules will improve HFSP fellows' training by giving them more time to adapt to a foreign environment and new equipment. "We will improve the repatriation rate of fellows. Q.S."

Japanese court backs woman researcher

Robert Triendl, Tokyo

For what is believed to be the first time, charges of sex discrimination against a professor in a public university have been upheld by a Japanese district court.

In a recent ruling, the government of Japan's Nara prefecture has been ordered to pay ¥550,000 (US\$5,000) in compensation to Kumiko Ogoshi, an associate professor in the department of public health at Nara Medical University.

In addition to some minor charges, such as the disposal of liquid waste in the plaintiff's office, the professor supervising Ogoshi was found guilty of "unjust distribution of research funds", "refusals to sign necessary documents", and "attempts to urge Ogoshi to resign or transfer".

But, referring to a legal exemption for government officials acting on duty, the court said that the compensation should be paid by Nara's prefectural government, rather than by the academic himself.

Three years ago, after more than two decades as an assistant professor, Ogoshi decided to sue her supervising professor on the grounds of unfair treatment and "bullying". She says that she took this action only when the situation became

"unbearable" after she launched an association of assistant professors in 1993, which drew attention to problems in Japan's academic research system. "I simply couldn't continue doing my research work any more," she says.

Many female scientists in Japan say they have been exposed to similar pressures. The percentage of female scientists in career positions is very small.

Only 7% of professors in Japan's public universities are female. And as most of these are concentrated in disciplines such as household economics or literature, the number in science and engineering departments is considerably lower.

This is in sharp contrast to an increasing number of female science graduates in the country — and a comparable increase in co-authorship by female scientists of highly cited papers. But, in a system where hiring and promotion decisions often depend heavily on personal contacts or recommendations from a senior professor, female scientists face significant career barriers.

A researcher at a national research laboratory who prefers to remain anonymous says her promotion to group

leader was evaluated negatively on the grounds of a lack of management experience. "But in reality, they just didn't want to have a woman at laboratory-chief meetings," she says.

In one high-profile case, Akiko Itai, an assistant professor in the pharmaceutical department at the University of Tokyo, decided to set up her own company after the university repeatedly refused to promote her to professor. Today she runs the Institute of Medicinal Molecule Design, a rational drug-design company and one of Japan's most successful new biotechnology ventures.

Ogoshi says she is pleased by the court's decision to award her compensation, but adds that the courts have shown "little understanding" for the situation in Japanese universities where an organizational model called the '*koza* system' gives professors almost unlimited control over their juniors.

Although Ogoshi says she does not expect the situation in Japanese universities to change soon, she argues that changes will never arrive "if women scientists don't speak out".

► <http://www.sorifu.go.jp/danjyo/index2.html>

Singapore to create nationwide disease database

David Cyranoski, Tokyo

Singapore's genetic diversity is set to come under scrutiny with the creation of a nationwide disease database. Containing both clinical records and genetic information, the project is the first major goal of the national genomics programme launched earlier this year.

The database is likely to combine personal data — such as medical history and family pedigree — with tissue, body fluid and DNA samples. Initially, it will cover patients suffering from common conditions such as cancer, heart disease, stroke and myopia. In the long term, the database will probably expand to cover health-related information on the general population.

The Singapore Genomics Programme (SGP) was set up in June with funding of S\$62 million (US\$35 million) over five years. Its original remit was to focus on novel genes and their related molecular targets in an effort to diagnose and treat diseases, such as liver cancer, that have a relatively high incidence in Singapore and Asia.

The idea of creating a broad-ranging database, rather than just focusing on a few specific diseases, is based on suggestions by the International Advisory Council (IAC),



Melting pot: Singapore's people are a mixture of Chinese, Malaysians and Indians.

which advises Singapore on life sciences. The council suggested that the SGP should take advantage of Singapore's unique diversity to establish a fundamental scientific resource.

Singapore's remarkably stable racial mix, featuring a large number of Indians and Malaysians as well as the Chinese majority, offers unique advantages compared with the homogeneity of other countries such as Iceland and Estonia. For example, researchers

will be able to follow the responses of different races to a given drug.

According to its director, Kong Hwai Loong, the SGP will now concentrate on building up the databases, manpower and core technologies — including microarrays and high-throughput sequencers — that will be needed for the broader study.

Based on another IAC recommendation, the SGP also plans to make its data freely available to all academic research institutes in Singapore. Kong acknowledges that many ethical issues, including informed consent and protection of privacy, need to be addressed. He notes that this will be especially important when the study is expanded to collect DNA and other samples from the 'normal' population — those who appear free of disease.

Much like the genetic study underway in Iceland, Kong anticipates that it will take a couple of years to sort out the ethical aspects of such a study. Although the SGP is not planning to examine a "major chunk of the population" as in Iceland, Kong acknowledges that study of the 'normal' population is inevitable. "It is just a matter of what level of sophistication we will be able to achieve," he says.

Canadian government combines tax cuts with science boost

Montreal As a prelude to a call for a general election, Canada's federal government has announced yet another round of tax cuts and extra spending for research, development and the 'new economy'. These follow closely on the Can\$58 billion (US\$38 billion) in tax cuts and more than Can\$4.1 billion in science spending included in last February's budget (see *Nature* **404**, 8; 2000).

This time, finance minister Paul Martin's 'mini-budget' includes an additional Can\$400 million for operating costs associated with capital investments for research funded by the Canada Foundation for Innovation, and an additional Can\$100 million to support infrastructure for Canadian participation in international research projects, to be decided by the foundation.

There will also be further tax cuts for middle- and high-income earners, businesses and investors. These are hoped to reduce the likelihood of research scientists and high-technology workers being attracted abroad, particularly to the United States.

A healthy fiscal surplus of Can\$6 billion, together with the emergence of a new federal

opposition party, the Canadian Alliance — which has adopted large tax cuts as one of its major election pledges — has convinced the reigning Liberals to call an early election next month, the third in seven years.

University of California will still manage weapons labs

San Diego The US Department of Energy announced last week that it will continue to contract with the University of California for the management of two beleaguered national weapons laboratories, the Los Alamos National Laboratory in New Mexico and the Lawrence Livermore National Laboratory in California.

The move follows several critical security and management reviews. The university has agreed to a number of management and security improvements. Its president, Richard Atkinson, says the department's decision to continue the contract "will be a welcome boost to our scientists and staff".

Centrifuge upgraded for earthquake work

San Diego One of the world's largest centrifuges, capable of spinning up to five tonnes, is undergoing a US\$4.6 million upgrade at the



University of California at Davis. The revamp will significantly expand its ability to do earthquake research.

Based at the university's Center for Geotechnical Modeling, the centrifuge (above) will be able to stress payloads of up to 80 times gravity, double the current gravity stress that can be tested at the facility. The National Science Foundation award for the improvements is the largest grant so far from the agency's US\$82 million Network for Earthquake Engineering Simulation programme.

UK polio vaccine withdrawn in response to BSE fears

London The British government last week withdrew stocks of a brand of polio vaccine that had been prepared using bovine serum. It was reacting to fears that the vaccine could be

contaminated with the agent that causes BSE. The move is in accordance with European guidelines which, although not yet legally enforceable, recommend that oral medicines should not be manufactured using bovine materials from BSE-affected countries.

The vaccine, manufactured by the pharmaceutical company Medeva, is used for routine vaccinations given to children and travellers. The government's chief medical officer Liam Donaldson said that the withdrawal was a precautionary measure. "I am advised that the risk of a person contracting vCJD from this oral polio vaccine is incalculably small," he said.

British physicist to lead ESA science

London David Southwood (above), professor of physics at Imperial College, London, and a former head of Earth Observation Strategy at the European Space Agency (ESA), has been appointed director of science at the agency. He succeeds Roger Bonnet, whose term of office has expired.

In his former role at ESA, Southwood set up and secured funding for the 'Living Earth' programme. He says that his main task in his new post will be to ensure the success of the ambitious package of new missions —



including a trip to Mercury and participation in the Next Generation Space Telescope — agreed at the meeting of the agency's council two weeks ago (see *Nature* 407, 277; 2000). "The community has spoken; my job is to deliver," he says.

Ontario refuses to cut greenhouse-gas emissions

Montreal Ontario, Canada's richest and most industrialized province, has refused to cut its emissions of greenhouse gases. The news comes a week after the country's federal government announced a plan — since agreed by all the other provinces — to take it one-third of the way to its Kyoto conference target in reducing emissions (see *Nature* 407, 824; 2000).

Ontario's environment minister, Dan Newman, challenged his counterparts to match his province's achievements in cutting emissions. But Ontario's inactivity in converting coal-fired power stations to cleaner fuels, as well as deep cuts to its environment ministry, have come under fire from environmental groups such as Greenpeace.

NIH funds training for Third World bioethics

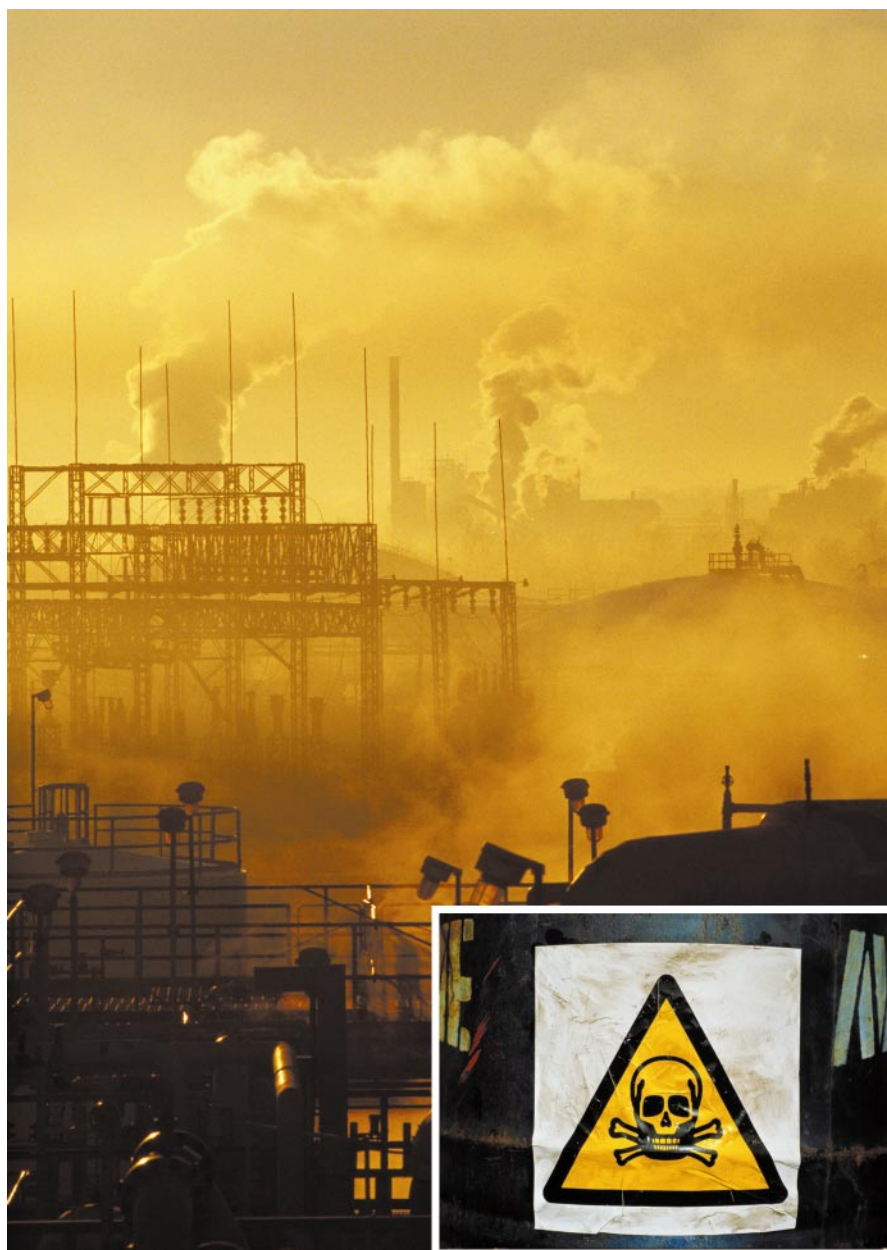
Washington The US National Institutes of Health's Fogarty International Center last week announced its International Bioethics Education and Career Development Award Program. The scheme will fund eight awards to support bioethics in developing nations.

The programme emerged from the Global Forum for Bioethics in Research, initiated by the centre and held in November 1999. Awards to train investigators from poorer nations in research ethics have gone to the University of Toronto, Johns Hopkins University, Baltimore, Albert Einstein College of Medicine, New York, Harvard School of Public Health, Cambridge, Massachusetts, and Case Western Reserve University, Cleveland. The universities of Cape Town, Pretoria and Chile received planning grants.

Correction The News Feature on drugs in sport (*Nature* 407, 124–127; 2000) described the new blood test for erythropoietin as the work of the Australian Institute of Sport. In fact, the project also involved the Australian Sports Drug Testing Laboratory, the China Doping Control Centre, the Norway University of Physical Education and Sport, and the universities of Quebec, Montpellier and Melbourne.

Clean and green...but are they mean?

Chemists are working to find alternatives to the noxious organic solvents that currently dominate their industry. As the leading candidates begin to hit the production plant, David Adam tests the atmosphere.



TONY STONE

TONY STONE

The chemical industry has an image problem. Despite progress in reducing pollution — and the efforts of corporate public-relations departments — belching chimneys, poisoned rivers and the risk of fire or explosions remain at the forefront of many people's minds. And as long as chemical plants continue to rely heavily on toxic and flammable organic solvents, these unsavoury images are unlikely to go away.

But techniques that eschew noxious organic solvents for more environmentally friendly alternatives are starting to emerge. Leading the way are two unusual materials. One is a hybrid phase of carbon dioxide, known as supercritical CO₂, which displays properties of both a liquid and a gas. The other is a group of ionic liquids — salts that remain molten at room temperature. Ionic liquids have already been successful in a commercial process that is now ready for licensing, and the world's first chemical plants using supercritical carbon dioxide as a reaction solvent should start up next year.

Green chemistry recently gained its own eponymous journal¹. But if this emerging

field is really to take off, supercritical CO₂, ionic liquids and their successors will have to prove more than their environmental credentials — the industry will demand practical benefits and economic viability. "Environmental benefits alone are never going to persuade people to spend money," says Martyn Poliakoff of the University of Nottingham, a leading figure in the development of chemical processes using supercritical CO₂.

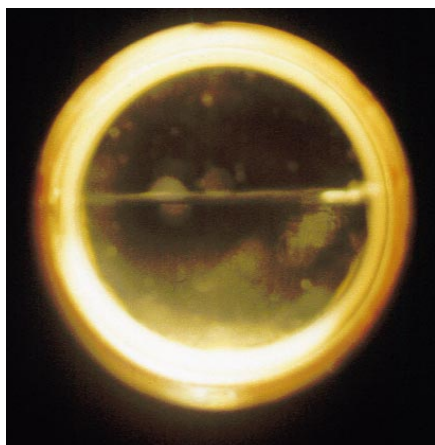
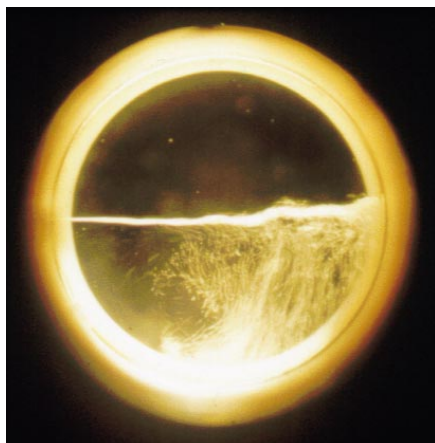
Help from a hybrid

Chemical reactions tend to be faster and easier to control when they take place in liquids because the reacting molecules can mix more readily. But many useful reactants are solids, and so must first be dissolved. Water can only dissolve a restricted range of substances, and it was the introduction of a range of organic solvents in the nineteenth century — including hexane and various alcohols — that opened the doors to thousands of new chemical processes. But many organic solvents are toxic and flammable, and usually they are very volatile. In addition, used organic solvents are difficult to recycle.

Chemists wanting to use different solvents have relatively few choices. Enter supercritical CO₂. Normally, when a gas is compressed it turns into a liquid, but above its critical temperature (31.1 °C) gaseous CO₂ does not. Instead, it enters a relatively dense, liquid-like phase that remains highly compressible and mixes easily with other gases. This makes it a very 'responsive' solvent: subtle changes in temperature or pressure can cause different chemicals to pop in and out of solution. Reducing the pressure switches the CO₂ back to being a gas, which is easy to recover and reuse.

Supercritical CO₂ is already in wide use for removing caffeine from coffee beans. But chemists admit that it has proved more difficult than expected to develop it as a reaction solvent. In fact, they started working towards that goal in the 1980s. "I think we oversold CO₂ back then," says Eric Beckman, a chemical engineer at the University of Pittsburgh. "People thought it would behave just like hexane and could be swapped straight in."

In reality, many key reagents do not dissolve readily in supercritical CO₂. But recent



Now you see it... The divide between liquid and gas blurs when carbon dioxide becomes supercritical. Under pressure (top) CO₂ is a liquid, but above both its critical temperature and critical pressure, it takes on remarkable characteristics that make it a prime candidate for green industrial chemistry.

breakthroughs have caused another surge in interest, prompted in part by a technology now being used for dry-cleaning. Four years ago, a team led by Joe DeSimone at the University of North Carolina at Chapel Hill discovered that a wider range of useful chemicals can be dissolved in CO₂ if they are tethered to certain fluorocarbons². Essentially acting as a detergent, these additives help to disperse dyes, catalysts and even water in CO₂.

In February 1999, Hangers Cleaners, a dry-cleaning firm based in Raleigh, North Carolina, used this technology in its first CO₂-based cleaning facility. Usually, clothes are dry-cleaned by being tumbled in the organic compound perchloroethylene. But this chemical is classed as an environmental hazard and can damage the nervous system, liver and kidneys. Using the fluorocarbon additives, grease and dirt are dissolved so effectively that the cleaning process can be carried out in liquid CO₂ — it is not even necessary to go supercritical. Hangers recently celebrated cleaning its millionth garment and plans to have 70 CO₂-based outlets by the end of the year.

Unfortunately, DeSimone's fluorinated



additives are too expensive to make using CO₂ economic in many industrial processes. But earlier this year, Beckman announced that his group had developed cheaper additives that should work in a wide range of situations³. The researchers combined polyether and polycarbonate polymers to create a long-chain copolymer that can smuggle insoluble reagents into CO₂. The flexibility of the molecule's skeleton and the strong attraction between CO₂ and the reactive groups on the polycarbonate make the copolymer — and other chemicals attached to it — dissolve more readily.

The new additives work better than the fluorocarbon detergents, Beckman says, and they are about a hundred times cheaper. Several companies have already bought licences for the technology. Although much of the ensuing work is cloaked in commercial secrecy, Beckman is prepared to reveal that the new additives have led to a cheaper synthesis of hydrogen peroxide.

Added value

Another important industrial reaction, hydrogenation — adding hydrogen to an organic molecule — is also being tackled using supercritical CO₂. Hydrogenation is one of the CO₂-based syntheses that does not need additives — in fact it is hydrogen's high solubility in supercritical CO₂ compared with organic solvents that makes this route so attractive. The British company Thomas Swan is investing £2 million (US\$2.9 million) in a pilot plant scheduled to open by April 2001 at Consett in north-east England. The plant will test a variety of hydrogenation processes using technology developed by Poliakoff at Nottingham⁴.

Poliakoff predicts that hydrogenation will become increasingly important as the chemical industry turns to biological molecules from plant materials, rather than oil, to make organic products. To make chemicals such as plastics from biological feedstock, the oxygen content must first be reduced by adding hydrogen. Supercritical fluids improve the yields of these reactions and make it easier to generate the desired prod-

ucts selectively, Poliakoff says. In addition, changing the fluid's solvent properties by altering the temperature and pressure allows the reactions to be tweaked to give a different range of products.

The chemicals giant DuPont, meanwhile, is building a US\$40 million development plant in Fayetteville, North Carolina, that will test the viability of supercritical CO₂ to produce the fluoropolymers used to make Teflon plastics. Teflon is usually made using chlorofluorocarbon solvents, but these are being phased out under the Montreal Protocol and finding suitable replacements has been problematic. The plant should open next spring and, if successful, DuPont says it will invest a further US\$235 million in the technology.

Meltdown

But supercritical CO₂ will not be suitable for all chemical processes. Many products have commercial values too low to justify the cost of switching plants from organic solvents. And some reactants and catalysts stubbornly refuse to dissolve in CO₂.

So green chemists are developing other options, of which ionic liquids — formerly known as low-temperature molten salts — seem to be the most promising. Usually salts are crystalline solids that melt only at high temperatures. Their strength lies in the interactions between the ions that make up the salt. Because the oppositely charged ions are small and similar in size, the attractive forces between them produce a strong structure that needs a lot of energy to pull it apart. But if this balance is upset by using an excessively large positive ion, the overall structure becomes looser, and the attractive forces within the molecule are weak enough to make the salt a liquid at room temperature.

The most common ionic liquids use organic salts such as 1-ethyl-3-methylimidazolium chloride to provide the large ion. These are combined with inorganic salts such as aluminium(III) chloride. When mixed together, these two solids produce an ionic liquid. By using different combinations of ions, and adjusting their relative concentrations, it is possible to make up solutions

▶ that dissolve a wide variety of substances including plastics, many metals and even some rocks. "For the first time it is possible to design a solvent to optimize a reaction rather than let the solvent dictate the course of the reaction," says Ken Seddon at Queen's University of Belfast, Northern Ireland, one of the leaders in the field.

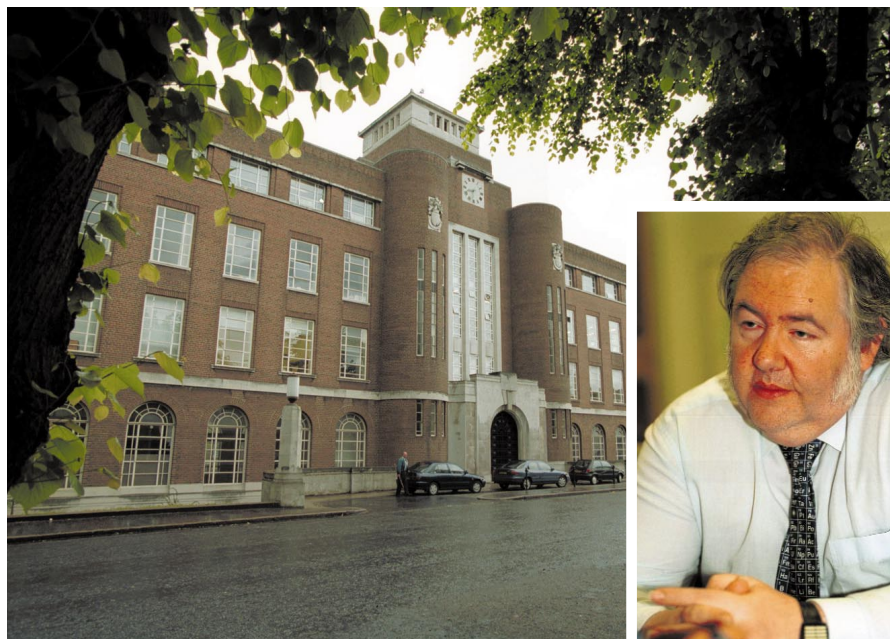
Here, too, the infant technology has taken its first steps towards industrial applications. A team at the French Petroleum Institute (IFP), an independent research centre near Paris, now has a commercial process based on ionic liquids ready for licensing⁵. The process exploits a major advantage of ionic liquids — the fact that they do not mix with many organic liquids. Chemists often want to dissolve catalysts but it is difficult and expensive to recover them from solution. But the IFP process uses a chloroaluminate ionic liquid to overcome this problem.

The reaction is a variation on the process for producing octene, which is used as a feed-stock for making phthalates — additives that soften many plastics. The IFP's conventional process produces around 200,000 tonnes of octene each year using butene as a starting material. But the expensive nickel catalyst needed is used as a solid, and is lost at the end of the reaction, says Hélène Olivier-Bourbigou, who leads the IFP group. In the new process the catalyst is dissolved in the ionic liquid and can be recovered and reused because the resulting octene does not mix with the chloroaluminate liquid.

There have been no takers for the new ionic liquid technology yet, Olivier-Bourbigou admits. But negotiations are continuing. "How fast ionic liquids could be introduced depends strongly on demonstrating real advantages for their use and whether they can be dropped in to existing processes," says Hans Meyer, research director with the pharmaceutical and chemical company Solvay Deutschland in Hannover. Olivier-Bourbigou argues that the IFP process performs well on both counts. "Sometimes it's difficult to make changes in our industry," she says. "We have demonstrated this technology on a pilot scale and shown that it could be applied to continuous industrial production."

Positive reaction

The IFP process is likely to be the first of several. Like teenagers stumbling across a crate of mixed spirits and a cocktail menu, chemists over the past few years have been experimenting eagerly with ionic liquids, ticking off the various reactions that the solvents permit. Already, various groups have recast important organic synthesis reactions with ionic liquids in the leading role. In the Heck reaction for instance, which links carbon atoms together, the palladium catalyst is usually lost after the reaction. By using an ionic liquid, Seddon's group can recycle it⁶. Other recent successes include Friedel-



Charged up: Seddon (inset) is pushing the field of ionic liquids forward at his lab in Belfast.

Crafts reactions⁷ (often used to make useful compounds from benzene) and the catalytic cracking of polyethylene⁸, which is an important way of recycling certain plastics. Normally, cracking creates a wide range of products, some of which present disposal problems. The ionic liquid process is more selective, generating a greater proportion of useful, low-molecular-weight products such as isobutane.

The nuclear industry is also interested: spent fuel could be dissolved in an ionic liquid, its components separated and some of them recycled. And recent results from a team of biochemical engineers led by Gary Lye at University College London show that ionic liquids could replace toluene as the solvent in the production of antibiotics⁹.

But Seddon believes the field is being held back by a lack of fundamental data on the characteristics of ionic liquids, such as their densities and viscosities. "It's like being a chemist in the nineteenth century," he says. To help gather that information, and share it among interested companies, Seddon and his Belfast colleague Jim Swindall last year established the Queen's University Ionic Liquid Laboratories (QUILL) consortium. Its members include industrial giants such as Merck, ICI and DuPont. Rather than competing to investigate the potential of ionic liquids, QUILL members share results and pool any patents. A seat at this round table costs each company £20,000 per year.

Others may soon have access to similar data. The US National Institute of Standards and Technology plans to extend its existing molten-salts database to include low-temperature ionic liquids. Access will be online and free. The QUILL data might also be released into the database, provided the consortium members give their approval.

Seddon believes that process engineers will eventually pick and choose ionic liquids or supercritical CO₂ for different applications. But research into the two leading green solvents is also converging. Beckman and fellow chemical engineer Joan Brennecke at the University of Notre Dame, Indiana, for example, have developed a process that uses supercritical CO₂ to extract reaction products from ionic liquids¹⁰. This could remove the need to find ionic liquids that automatically spit out the product they help make. Brennecke's team has already managed to remove naphthalene from an ionic liquid known as [bmim]PF₆, and she believes supercritical CO₂ could be used to scrub clean most industrial ionic liquids.

But nobody in the field of green solvents is overselling the technology. Thomas Swan will know within about a year if its supercritical CO₂ hydrogenation plant is successful. The company's director and general manager, Dai Hayward, although optimistic, stresses that the process — like many based on the new solvents — has yet to prove itself economically. If it cannot compete then there will be little room for sentiment. "We're willing to lead with our chin on this," he says. "But there's only so far we're going to stick it out."

David Adam is a member of Nature's science writing team.

1. Leitner, W. *Nature* **407**, 295–296 (2000).
2. McClain, J. *et al. Science* **274**, 2049–2052 (1996).
3. Sarbu, T., Styranc, T. & Beckman, E. *Nature* **405**, 165–168 (2000).
4. Hitzler, M. G. & Poliakov, M. *J. Chem. Soc. Chem. Commun.* 1667–1669 (1997).
5. Olivier-Bourbigou, H., Chodorge, J. A. & Travers, P. *Petrol. Technol. Q.* **4**, 141–149 (1999).
6. Carmichael, A. *et al. Org. Lett.* **1**, 997–1000 (1999).
7. Adams, C., Earle, M., Roberts, G. & Seddon, K. *J. Chem. Soc. Chem. Commun.* 2097–2098 (1998).
8. Adams, C., Earle, M. & Seddon, K. *Green Chem.* **2**, 21–24 (2000).
9. Cull, S. *et al. Biotechnol. Bioeng.* **69**, 227–233 (2000).
10. Blanchard, L., Hancu, D., Beckman, E. & Brennecke, J. *Nature* **399**, 28–29 (1999).

Disease should be targeted from all angles

Sir—In his review¹ of my book, *Cancer: The Evolutionary Legacy*, Leo Kinlen is unpersuaded that an evolutionary perspective on cancer causation is helpful. As an epidemiologist, Kinlen is more than entitled to hold contrary views to me. Scientists of differing specialties often have quite different perspectives on causation². They each know what they mean, but a broader audience may be perplexed by limited horizons. That aside, Kinlen's critique raises substantive issues that merit further comment or debate.

On the contentious issue of causal mechanisms in breast and prostate cancer, Kinlen has, unfortunately, misconstrued my views. He says that I propose that environmental chemicals *must* (my emphasis) make a significant contribution to the cause of breast cancer and that sex is a *likely* cause of prostate cancer. On the contrary, on breast cancer I argue that the theory that environmental chemicals are an important cause of breast cancer has virtually no epidemiological support. I do say, however, that since unequivocal negation of the hypothesis is very difficult, it cannot be completely excluded. On prostate cancer, I state that I do not know the cause and that the idea suggested, though plausible, is "highly speculative".

On the challenge of cancer control, Kinlen reminds us that epidemiological insight into cancer causation and effective prevention is sometimes possible without any detailed knowledge of biological mechanisms, well-known examples being cigarette smoking and lung cancer, and melanoma and sun exposure. By the same token, childhood leukaemia was curable before we knew much about the biology of that disease. These epidemiological and clinical successes are laudable, but they also bring into sharp focus our past limitations and failures. Would cigarette manufacturers have been able to prevaricate for so long if it had been clearer that particular chemical carcinogens in cigarette tar mutate *p53* and other genes, and if the long latency (decades) and less than odds-on risk (about 1 in 10) of lung cancer in moderately heavy and persistent smokers had a persuasive molecular genetic explanation?

Traditional epidemiological approaches have not been able to deliver clear verdicts on causation for many of the other major cancer types—breast, prostate and colo-rectal carcinoma in particular. Similarly, success in the treatment of childhood leukaemia and a few other cancers has not yet been followed by major therapeutic inroads into the clinical

control of adult epithelial carcinomas.

We and the general public are entitled to ask, why not? Is it not due in some measure to the fact that, until quite recently, we have pursued our clinical and epidemiological businesses in ignorance of the underlying biological complexity? But then, as Kinlen might reply, a 'laboratory scientist' would say that, wouldn't he?

M. F. Greaves

Leukaemia Research Fund Centre, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

1. Kinlen, L. *Nature* **405**, 122–123 (2000).

2. Rose, S. *Lifelines* (Alan Lane/Penguin, London, 1997).

Returners not welcome at Spanish universities

Sir—*Nature* has published several articles regarding some of the obstacles that the Spanish scientific community faces in matters of appointments and research funding policies (*Nature* **396**, 709 & 712 (1998); **400**, 203 (1999); **405**, 723 (2000)).

We would like to draw attention to the case of many researchers who, having completed doctoral or postdoctoral training abroad, return to Spain with the financial support of very selective programmes sponsored by the EU, the state or the regional governments. The number of these "reincorporation contracts" is very small, and they are awarded after meticulous selection on the basis of scientific achievement. However, the selected candidates have to face discrimination in several areas in order to continue their academic careers.

As a group of current and former reincorporated researchers at the Universitat Autònoma de Barcelona (UAB), we believe our experiences are typical of the difficulties faced by people in our position.

Despite long negotiations with the university rectoral team, we still do not enjoy full academic status, our teaching and scientific activities at the UAB notwithstanding. We do not belong to the university elected body, where academic and administrative personnel as well as students are represented. Consequently, we do not have democratic rights. This situation is common to many Spanish universities.

Most importantly, the UAB officials have designed a lecturing policy document—now being implemented—that involves the creation of assistant professor positions for which only current UAB assistant lecturers are eligible. This allows our own PhD students to apply for these posts, while neither UAB researchers like us nor those from outside the UAB are eligible even if they are more qualified.

Our university was recently named as the top-scoring institution in a pioneering study, sponsored by the Ministry of Education and Culture, on the quality of Spanish universities in terms of educational development, organization, teaching resources, women's participation, doctorates and student success (*Nature* **402**, 848; 1999). Yet we are still subject to blatant endogamic practices, a barrier to the further development of high-quality research and teaching in the university.

Pau Ferrer

Department of Chemical Engineering, Universitat Autònoma de Barcelona, 08193-Bellaterra, Spain

Other signatories to this letter:

Anna Gavarró, Josep M. Palet, Jordi Surrallés, Lluís M.

Mir-Martínez, Jordi Ortiz, Alexandra Uscatescu, Lydia

Giménez-Llort, Teresa Ferriz-Roure, Begonya Saez-Tajafuerce

UAB, Bellaterra, Spain

Josep Quer University of Amsterdam, Amsterdam, The Netherlands

Maria A. Escobar ESEPR SA, Barcelona, Spain

Sandra Montón-Subías Universitat Pompeu Fabra, Barcelona, Spain

Carles X. Simó Universität Bielefeld, Bielefeld, Germany

Ramon Herrero Universitat Politècnica de Catalunya, Barcelona, Spain

No evidence to link polio vaccine with HIV ...

Sir—We would like to add two clarifications to your editorial and News story about the Royal Society meeting on the origin of human immunodeficiency virus (HIV) and the hypothesis that oral polio vaccine may have been contaminated^{1,2}.

First, contrary to your statement, and as presented at the meeting, only two batches of vaccine were used in the mass vaccinations conducted in the Congo before 1959, the date of the first confirmed HIV infection in a human. No concrete or credible evidence exists that polio vaccine was produced locally in the Congo, and neither of the large lots produced in Philadelphia have been found to contain simian immunodeficiency virus (SIV) or HIV by the very sensitive polymerase chain reaction test conducted by independent laboratories³.

Second, you tax people in some other fields, particularly those working on xenotransplantation, with complacency about cross-species contamination. In the case of polio, we adapted attenuated vaccine strains to human diploid strains as soon as they became available, precisely because of concern regarding extraneous agents⁴.

Stanley A. Plotkin*, Hilary Koprowski†

**Emeritus Professor, University of Pennsylvania, Doylestown, Pennsylvania 18901, USA*

†Thomas Jefferson University, Room M85, Jefferson Alumni Hall, Philadelphia, Pennsylvania 19107, USA

1. *Nature* **407**, 115 (2000).

2. Dickson, D. *Nature* **407**, 117 (2000).

3. Plotkin, S. A. *Phil. Trans. R. Soc. Lond.* (in the press).

4. Hayflick, L., Plotkin, S. A., Norton, T. W. & Koprowski, H. *Am. J. Hyg.* **75**, 240–258 (1962).

Sexually transmitted retribution

A history of attitudes towards sinning and its consequences.

The Wages of Sin: Sex and Disease, Past and Present

by Peter Lewis Allen

University of Chicago Press: 2000. 202 pp.

\$25, £17.50

W. F. Bynum

Have you noticed how many dust-jackets these days carry recommendations from Roy Porter? Four of the five books I took to the Canaries last November carried the Porter Seal of Approval. Two of the three books I am using in a course I am teaching this autumn have Porter puffs; he edited the third one.

The dust-jacket of Peter Allen's new book has the following by Porter: "Quite the best survey there is of the sad history of love and disease, sex and sin. Peter Allen writes with panache and compassion." Well, yes and no. This series of case studies of historical attitudes towards the relationship between disease and sin is a good read. It is engaging and intensely personal, with a moving autobiographical introduction that identifies the author as a gay Jewish male one of whose earliest lovers died in his arms from AIDS.

Allen also attempts to see both sides of his story. Christianity has been instrumental in defining sin within Western culture for the past two millennia; most of those whom Allen singles out as eager to cast first stones, or to associate disease with sin and sufferers with just punishment, were certain that God was on their side. Christianity is not the only religion to castigate the sex-for-fun brigade, but for those Allen writes about, and writes for, it is the dominant moulder of values, especially within the born-again United States. (An angry Muslim, Hindu or Jew could have told much the same story as Allen, within different cultural contexts.)

Most of Allen's bad guys are thus Christians. But so are most of his good guys and women. While highlighting the religious origins of prudery and fear of the body and its sexual functions, he also recognizes that Christianity can be a religion of compassion and healing. He appreciates the religious origins of hospitals and the ministrations of those who cared for sufferers of leprosy, syphilis and bubonic plague in times past, and who minister to people with AIDS today. He even has sympathy for Saint Augustine, one of the key players in the evolution of negative Christian attitudes towards the human body, perhaps because even Augustine could famously pray for chastity and continence, but "not just yet".

The dichotomy that Allen seeks to analyse is most subtly exposed in the first (and best) chapter, on lovesickness. This disorder,



Pleasure before pain: Hieronymus Bosch's *Garden of Delights* (1503–04).

clearly described by the Greeks, could affect men and women of all ages, though of course it was more common before old age. Caused, among other things, by unrequited love or a fixation on some love object, it led to listlessness, loss of appetite, melancholy and, sometimes, death. For the Greeks, the treatment was unproblematic: sex. For medieval Christians, steeped in a culture that agreed with Saint Paul that marriage is only marginally better than burning, and sex outside marriage a deadly sin, what was a doctor to prescribe after diagnosing lovesickness? If his patient was unmarried, marriage might work. But what if his patient was already married and the source of this ailment was not his or her spouse? A few hardy doctors argued for following Greek precedent, but theologians thundered against the very notion of coition as cure.

After lovesickness, Allen offers us the standard group of diseases that, for a variety of historical reasons, have been singled out for stigmatization: leprosy ('the disease of the soul') and plague, for example, which were often associated with non-sexual sin. But for the most part Allen is interested in the consequences of sexual indiscretion, including an interlude on the consequences of masturbation. He has a good eye for the occasional novel quotation, but also a wonderful tendency to rediscover the wheel at frequent intervals along the way.

There are two substantial problems with this book. One is its very familiarity. Allen's central theme was dredged up repeatedly in the early 1980s, when AIDS was new. Maybe it does all need saying again, but I thought we

were pretty familiar with the 'diagnosis as moral judgement' motif, with homosexuality medicalized and then publicly demedicalized in the past couple of decades. And what is 'lifestyle' medicine if not moral medicine? There is pretty good epidemiological evidence that many of the seven deadly sins can cause disease. (The nicest one is probably sloth, but even that can lead to cardiovascular problems.)

The second problem with Allen's book is that it is simply too much about the present. All history reflects the writer's present, and some historians (and sociologists) go so far as to suggest that history is only about the present. This book would be grist to their mills. Allen makes too many specious and intrusive comparisons between plague-ridden London and AIDS-ridden San Francisco, between medieval churchmen ostracizing lepers and Ronald Reagan refusing to mobilize American medical might against AIDS. Both were lamentable, but historians need to do more than simply draw shrill parallels.

For all of its admirable qualities, then, Allen's book doesn't stand up very well as history. I think he would probably be a witty and charming dinner companion, however. So, just as we are supposed to hate sin but love the sinner, we can value what he tells us about the contemporary United States even while he is writing about syphilis in sixteenth-century Italy.

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.

..... Ways to get them through the door

Museums of Modern Science (Nobel Symposium 112)

edited by Svante Lindqvist
*Science History Publications: 2000. 216 pp.
\$39.95 (hbk), \$24.95 (pbk)*

Graham Farnelo

The announcement of the science Nobel prizes a few weeks ago was covered well by the media, but not by science museums. As usual, the museums' reaction was not to stage a quick-response exhibition, but instead to indulge in their annual flurry of concern. Embarrassed by the paucity of their contemporary coverage, many of these museums — and their sibling science centres — felt obliged to say something coherent about the Nobel news. The result was usually a sheaf of downloaded web pages or pinned-up newspaper clippings that museum visitors all but ignore.

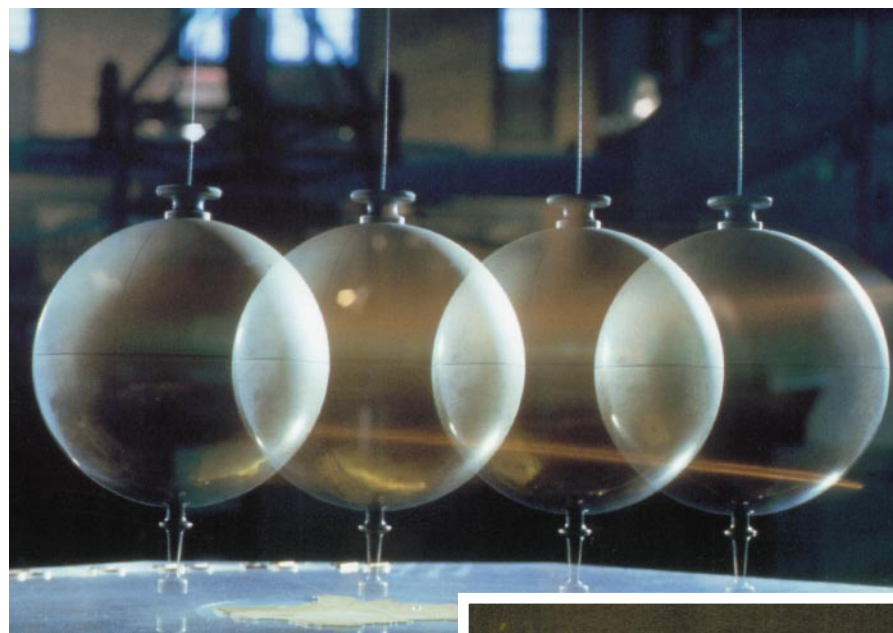
Yet whenever visitors are asked, they say they want more contemporary science. The onus is on museums to present this science in ways that make effective use of their great strength: the ability to present what physicist and Barcelona science museum director Jorge Wagensberg has memorably dubbed "the unrenounceable priority of the real" — special things and 'live' events.

But it's hard to bring these virtues to bear on most fields of contemporary science and technology. Most of the 'special things' are usually either too small, too big, too expensive or simply too boring to make attractive display material. Where are the visually arresting icons of xenotransplantation and nanotechnology?

Museums of Modern Science is a collection of essays that explore the issues involved in presenting contemporary science to museum audiences. The pieces are records of presentations given at a Nobel symposium last year in Stockholm, held partly to inform plans for the proposed Nobel Museum.

Such a museum could be a great cultural asset for science and the Swedish capital, once the developers have secured a building and funding. They also need items to display; at present they have only one, although it is a humdinger — the original will of Alfred Nobel, dated Paris, 27 November 1895. Other relevant artefacts ('objects', curators call them) are to be found in other museums, universities and research centres, usually in displays relating to in-house Nobel laureates.

One of the first decisions the Nobel Museum's planners must make is to choose their prospective audience. Is it the dignitaries at the annual ceremony? Is it visiting academics? Interested journalists? The lay public? School parties? As Alan Friedman of



Museum past and future: Foucault's pendulum (above) and Bob Millar's *Sun Painting*.

the New York Hall of Science points out in his telling essay, it is extremely difficult to keep all the possible visitor constituencies happy, let alone the funders. In the end, museums have to strike a balance. The challenge for every director is to decide on a compromise based on reasoned arguments and on analyses of visitors' needs and wants, not on which lobby shouts loudest.

But the toughest challenge for a science museum is to present its collections in ways that are both authoritative and engaging. The Cambridge historian Simon Schaffer points out the difficulty of interpreting recent history when its significance is still up for grabs. These days, museum exhibition developers usually arrange their objects in the context of a story chosen to appeal to visitors. The Smithsonian Institution's Bernard Finn complains that this has led to "an arrogance of ideas", with the objects becoming subservient to the exhibition theme.

The question of who should decide the content of exhibitions is perennially controversial. Should it be curators, historians, scientists or the marketing department? Whoever it is should beware of consulting too narrowly, the historian John Heilbron argues. He points to the exhibition that the Smithsonian's Air and Space Museum planned in 1995 about *Enola Gay*, the plane that carried the atomic bomb to Hiroshima.

The first draft of the exhibition made Japan appear as the primary victim of the Second World War. The resulting vociferous outrage of war veterans and influential politicians on Capitol Hill led to the dismissal of the museum's director. But the mistake made by the exhibition's curators, Heilbron argues persuasively, was that they consulted historians who had similar views, taking insufficient



account of other interested parties. Historians can freely interpret events of long ago, but woe betide them if they try to impose their opinions on events for which the participants are still alive. No one has a monopoly on interpreting recent history.

Heilbron also has cautionary words for his fellow contributors, especially the Wellcome Trust's Ken Arnold and Wolf Peter Fehlhammer, director of the Deutsches Museum in Munich, who advocate the deployment of contemporary art in science exhibitions. Clearly, such works add valuable context to other parts of a display and can, by virtue of their very presence, appeal to new audiences. But as Heilbron observes, the inclusion of art brings substantial additional open-endedness to a science exhibition. For it is surely "even harder to control responses to works of art than to objects of science".

The essays in *Museums and Modern Science* comprise a valuable manual of key strategic issues for the developers of the proto-Nobel Museum to ponder. The rest of the museum world will also benefit and should be grateful to the Nobel Museum's

staff for producing such a handsome and stimulating volume. But will this gratitude extend to lending the new museum their Nobel objects? ■

Graham Farmelo is at the Science Museum, London SW7 2DD, UK.

Life and times of a pioneering polymath

George Perkins Marsh: Prophet of Conservation

by David Lowenthal

University of Washington Press: 2000. 650 pp. \$40, £26.95

Thomas R. Dunlap

George Perkins Marsh is not a household name, except perhaps in the households of environmental historians. There he is remembered for *Man and Nature* (1864), the first study of humans as a worldwide agent of geological and biological change. In nineteenth-century America he was moderately well known in higher political, social and intellectual circles. He was a scholar, lawyer, congressman and for two decades a US diplomat in the Levant and the Italian states. David Lowenthal's book, which started as a revision of his earlier biography, *George Perkins Marsh: Versatile Vermonter* (1958), blossomed into a fresh attempt to understand Marsh and his influence.

Lowenthal is perhaps the ideal biographer for Marsh. He is a geographer, and so shares Marsh's environmental interests. And he has been interested in Marsh for many years. This interest sustained him as he tracked his man through libraries and archives and followed in his footsteps across various landscapes.

Knowledge and enthusiasm are necessary, for formidable obstacles are placed in the biographer's path by the very diversity of Marsh's life. He was a polymath. He knew many languages and wrote on a wide variety of topics. His interests included the English language and several others, philology in general, agriculture, soil erosion, the influence of the Goths on European civilization, political economy, public affairs and camels.

Marsh's public career was equally diverse. He made one of the first studies of the environmental effects of the fishing industry and forest clearance, investigated Vermont's railroad companies, designed the Vermont State House, represented that state in Congress, and served as an American consul and minister. He was a lawyer, invested in railroads, and was partner in a marble quarry — among other ventures.

The obvious approach for a biography would be to focus on what Marsh is remembered for, his ideas about humans and their

relation to nature. The book's subtitle suggests this tack, but the book is actually a more ambitious 'life-and-times'.

There are, in fact, two books here. One is the story of an American scholar, the other the record of Marsh as a prophet of environmental thought. Both are worthwhile, but for different reasons. The first puts Marsh in the context of his times. It was quite a context. Marsh lived through American politics from the anti-Masonic agitation around 1830 through to the end of Reconstruction (although he was out of the country for many of the later events). He saw American business expand from the shaky financing of the early nineteenth-century railroads to the industrial progress after the Civil War. He was involved in American scholarship and science in ways as different as the political manoeuvring that established the Smithsonian Institution and the 'dictionary battles' over American versus English usage and spelling.

Then there were the political events in Europe and the Ottoman Empire, the background to Marsh's diplomatic efforts. In addition, Lowenthal describes Marsh's family life, his travels, financial troubles and political problems: it is easy to see how Marsh's career was related to events. He could not have written with authority on so many topics even a generation later. His career as diplomat and wandering scholar was equally a product of American and European involvement in the countries where he was stationed. The profusion of events and ideas in the book and the detail on Marsh's travels and studies will discourage some readers, but the result is a vivid portrait of Marsh against his intellectual and social background.

The second story is the more important. As Lowenthal admits, "it is above all for his crucial role in environmental history that Marsh's life warranted retelling here". Most of that is described in the two chapters in which Lowenthal discusses *Man and Nature* and the final one on Marsh's ideas. But the environmental aspect runs throughout the text, for Lowenthal believes Marsh's circumstances and experiences shaped his ideas on conservation. He had seen the forests of his native Vermont fall to the axe and travelled over the eroded landscapes of the Middle East.

Experiences such as these, as much as formal knowledge of geology, made *Man and Nature* what it was. Lowenthal's assessments are judicious, usually more solid than startling, but occasionally bold. His comparison of Marsh and his contemporary Henry Thoreau should stir some thought. He rejects as a "latter-day construct" the idea that there was a division between biocentric and anthropocentric views, as represented respectively by Thoreau and Marsh, and finds many points of agreement between the two men. Then there is his view that, next "to Darwin's *On the Origin of Species*, Marsh's *Man and Nature* was the most influential text of its time to link



Advance guard: Marsh was a century ahead of his time in recognizing environmental concerns.

culture with nature, science with society, landscape with history". Well, maybe, but wasn't second place in this race a long way back?

There is something here for everyone. General readers will get a picture of Marsh and his times, American and European. Environmental historians and geographers will appreciate Lowenthal's discussion of Marsh as an environmental thinker, the fate of his ideas since 1864, and their relevance today. Everyone should come away with a better appreciation of a man who was a century ahead of the general run of scientists in recognizing many of our environmental concerns and who addressed them at a fundamental level. This is a useful study of an important figure. ■

Thomas R. Dunlap is in the Department of History, Texas A & M University, College Station, Texas 77843-4236, USA.

Parasites at the heart of ecology

The Spatial and Temporal Dynamics of Host-Parasitoid Interactions

by M. P. Hassell

Oxford University Press: 2000. 208 pp. £39.50, \$70 (hbk); £19.95, \$34.95 (pbk)

Ted J. Case

About 10% of the world's animal species make their living by laying eggs inside another insect or arthropod. This parasitoid lifestyle presents challenging problems for host and parasite alike. The parasite must avoid the immune responses of the host and synchronize its life history and habitat selection to match the host. It must have dispersal powers

to find host individuals but stay small enough to fit, as a juvenile, inside the host's body. The host must be reproductively prolific enough to cope with suites of such parasites taking their share of its reproductive output.

Simple models in ecology often take the form of a pair of coupled dynamic equations; one equation for each of the interacting species. Yet species interactions in nature are often much less species-specific. This tends to create a mismatch between our simple models for pairs of interacting species and the more web-like interaction networks that are the norm. For host-parasitoid interactions, this mismatch between theory and reality is less extreme. It is typical for a parasitoid species to be restricted to a single host species. And, for the most part, a single egg laid in a host translates into one less survivor for the host population and one more recruit to the parasite population. Host-parasitoid systems are thus the ground zero for theoretical ecology. If we can't get it right for these specialized interactions, we are probably not going to get it right anywhere.

There is another very practical reason why host-parasitoid interactions are central to ecology and why this book should be broadly read. The specialization of these species makes them prime candidates for use as biological control agents. The more specialized the predator or parasite, the fewer non-target side effects. Most insect pests are exotic species to an area, introduced by accident. Once freed from their native enemies, they can grow to pest levels. The challenge is to find a suitable parasitoid in their homeland. It must be deadly enough to reduce the pest populations radically, but not so efficient that they crash too low to support the parasite population, since that would return us to our starting point. We'd like to be able to identify characteristics that will lead to stable control at low host levels for each particular pest.

It would be hard to think of anyone better suited than Michael Hassell to review this field. His new book is a well-organized compendium of the myriad features that make or detract from stability in these tight interactions. Hassell last summarized the field in a monograph published in 1978. Since then, there have been major new developments.

Hassell details the important role spatial heterogeneity plays in coexistence and control. Also new is a growing list of theoretical studies that include webs of interactions among several host-parasite combinations. He is careful throughout to point out current deficiencies in both our theoretical and empirical understanding of these systems.

This book is a must-have for anyone interested in the theory of host-parasite interactions, and for those who just want to know more about ecological dynamics. ■

Ted J. Case is in the Department of Biology, University of California at San Diego, La Jolla, California 92093-0116, USA.

Science in culture

Colouring it true

Origins of the art of colour reproduction

Philip Ball

Art is most often viewed at one remove: as a reproduction in a book or on a poster. Usually without acknowledging it, we entrust our experience of the colours of Renaissance Venice or Impressionistic Paris to the skill and diligence of the printer. But a comparison of the same image in different books is often a sober reminder of the vagaries of colour reproduction.

Capturing colour on the printed page is one of the themes of "More Than Meets The Eye", an exhibition at the Victoria and Albert Museum in London that explores the science in art and design, ending on 3 November (see *Nature* **407**, 20; 2000). As this part of the exhibition shows, a knowledge of colour theory is of only limited help in overcoming the infidelities imposed by shortcomings in the technology and materials of printing.

Printing in many colours did not become commonplace until the nineteenth century. Some of the most glorious colour prints of this period were a technical *tour de force*, for each individual colour was typically applied by a separate printing plate. William Savage, appointed by the Royal Institution in London to improve printing technology, laboured for eight years on an illustrated book, *Practical Hints on Decorative Printing*, finally published in 1823, in which some of the images bore the imprint of no fewer than 29 separate woodblocks.

But a technique that was in principle more economical of materials and labour had been developed 100 years earlier. By the start of the eighteenth century artists and scientists had reached a consensus that there were but three primary colours, as well as the white and black needed to lighten or darken them. Said Robert Boyle in 1664: "There are but few Simple and Primary Colours (if I may so call them) from whose various compositions all the rest do as it were Result ... I have not yet found, that to exhibit this strange Variety [painters] need employ any more than *White and Black, and Red, and Blew, and Yellow.*"

To the French artist and engraver Jacob Christoph Le Blon (1667–1741), this suggested a way to create full-colour prints using just the three primary inks. If they were translucent, their superposition could generate the secondary colours (orange, green, purple), as well as tertiaries and more complex shades. Black, thought Le Blon, should arise from superimposing red, yellow and blue.

To capture tonal variations, Le Blon used the half-tone technique of mezzotinting. A metal plate was burred all over with a sharp implement, and then smoothed back down by hand to a degree proportional to the lightness of the image: smoother areas retained less colour when inked. But to prepare the three 'colour separation' plates



A colour print by Le Blon, from around 1722.

in the pre-photographic era, Le Blon had to pull off the astonishing feat of decomposing a full-colour image into the three primaries by eye.

He began to use this method in the early 1700s, but failed to find a sponsor until he came to Britain in 1719. Here, in collaboration with the wealthy dignitary Colonel Sir John Guise, he set up a company called The Picture Office in 1720. With the permission of King George I, the partners made several thousand copies of pictures from Kensington Palace.

They were impressive, by some accounts. Sir James Percival said of one of Le Blon's prints in 1721, "Our modern painters can't come near it with their colours, and if they attempt a copy make us pay as many guineas as we now give shillings." This, however, was the opinion of someone unused to seeing reproductions in anything but monochrome. In reality the method had several shortcomings. Because the inks were not pure primaries, their mixtures produced somewhat dirty colours — which time has only muddled further. The three primaries did not mix to black but to murky brown, so Le Blon was forced to add black laboriously by hand. And the plates lost their crispness after many impressions.

Le Blon's biggest handicap, however, was a poor business sense. The writer Horace Walpole considered him "either a dupe or a cheat, I think the former". Forced to flee England to escape his debts, he died in poverty. His three-colour process was abandoned until photolithography, combined with James Clerk Maxwell's invention of colour photography, made it practical in the 1860s. ■

Philip Ball is a consultant editor at *Nature*.

"not science AND art", a closing talk for the "More Than Meets The Eye" exhibition, will be given by art historian Martin Kemp at the Victoria and Albert Museum on 3 November (7 pm). Entry free.

High achiever

The discovery of the stratosphere laid the foundations of geophysics.

Mott T. Greene

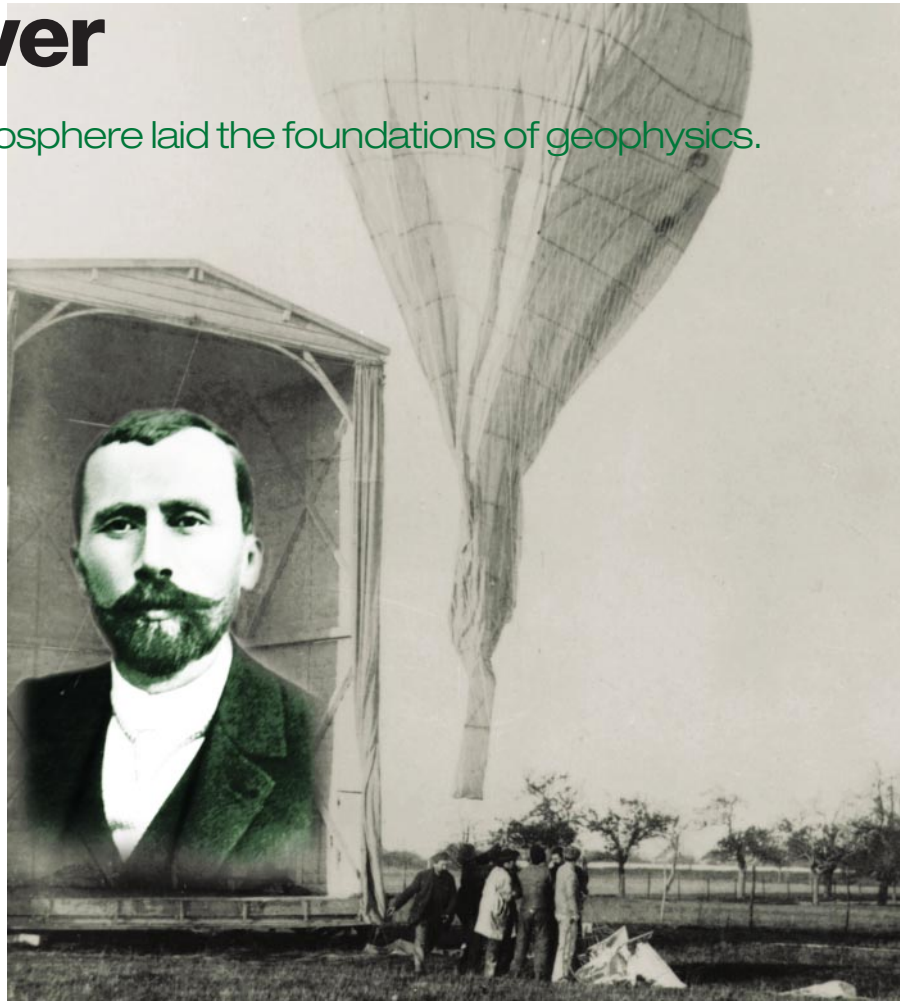
Between 1900 and 1905 the theories of quantum mechanics, special relativity and radioactive transmutation were born, and the atomic theory of matter was confirmed. These events so dominate historical accounts of the period that other breakthroughs languish in obscurity, among them the discovery of the stratosphere in 1902 by the independent French scientist Léon Teisserenc de Bort (1855–1913) — one of the most overlooked events in the history of science.

Teisserenc de Bort, working at his private estate at Trappes, near Paris, was a pioneering investigator of what were then the upper reaches of the atmosphere, up to altitudes of 14 km. He used balloons of his own design, delicate things made of kerosened paper and filled with hydrogen, trailing packets of recording instruments — also of his own design — which parachuted back to Earth.

In 1899 he began a new series of experiments. He had caught the public eye that year when one of his large meteorological kites had broken away, trailing 7 km of wire, and had cut the telegraph link between Paris and Rennes on the day the whole world awaited the verdict in the Dreyfus case. He therefore abandoned kites for a while, and returned to his upper-air work with balloons.

The temperature recordings from these new balloon flights were odd. Temperatures decreased with altitude by about 6 °C per kilometre, as theory predicted, but at 11 km or so the temperature record stabilized, usually somewhere between –52 and –54° C, and remained constant for several kilometres. At first, it looked like instrument error — perhaps solar heating of the instrument housing, re-radiated to the thermometer. Teisserenc de Bort wrapped his instrument housings in cork, but saw the same result. He moved the thermometer (a curved strip of two dissimilar metals, expanding and contracting at different rates) outside the housing. Still, the temperatures remained constant, and occasionally even rose. He began to fly the balloons at night to eliminate solar radiation. No change. He decided there might be a seasonal effect and tried flights throughout the year. The signature persisted.

He worked for two full years on the problem, and in 1902, having accumulated 236 ascents to buttress his argument, asserted the existence of an ‘isothermal zone’ of varying thickness, where the rate at which temperature decreased with altitude diminished to zero, usually at an altitude of 11 km, after



Teisserenc de Bort (inset) was puzzled by what his balloons told him about the upper atmosphere.

which the temperature was constant for several kilometres. The layer was higher over the high-pressure centres known as anticyclones, and lower over low-pressure cyclones. In consequence, he suggested, the problem of the general circulation of the atmosphere would have to be reconsidered, as such an isothermal layer meant that the atmosphere as a whole was not in convective equilibrium, but only that part bounded above by this isothermal zone.

In Germany, Richard Assmann, a much better funded, supported and equipped government meteorologist, read of the work and realized (but too late!) that he had seen the same phenomenon, but had put it down to instrument error. He reviewed the records of his own balloon flights and found that the temperature typically increased for a while above 11 km: Teisserenc de Bort’s isothermal zone was really a permanent temperature ‘inversion zone’. In the end, neither name stuck, although Teisserenc de Bort came up with names that did stick: the troposphere — ‘sphere of change’ — and above it the ‘stratosphere’, a sphere of constant laminar wind-flow around the planet.

Ripples from the discovery spread far beyond meteorology. Between 1902 and 1904

the Swedish oceanographer Vagn Ekman looked for, and found, similar layering of the ocean. In 1909, Andrija Mohorovičić, a Croatian meteorologist, used seismology to establish the existence of a similar discontinuity in the solid Earth now called the ‘Moho’ or Mohorovičić discontinuity.

The discovery that the Earth–ocean–atmosphere system is composed of concentric shells of different density, from the core of the Earth to the top of the atmosphere, is the founding insight of modern geophysics. This discovery also profoundly influenced the thinking of a young meteorologist named Alfred Wegener, leading him in 1912 to propose the theory of continental drift, with the continents representing the remains of a formerly continuous Earth shell above the ocean floors. Just as air masses and ocean water masses moved under the influence of the Earth’s rotation, sliding along surfaces of discontinuity, so, he reasoned, did the continents on a longer timescale — making Teisserenc de Bort not only the discoverer of the stratosphere but an honorary grandfather of continental drift. ■

Mott T. Greene is in the History Department, University of Puget Sound, Tacoma, Washington 98416, USA.

A Leap of Faith.

BY PROF. THEO VON HOHENHEIM.

IT is hard to say whether the speculation over the so-called Perry–Dean Shift will be exacerbated or alleviated by the results presented by Professor Ilan Goethe (the great poet’s nephew) in a recent issue of *Zeitschrift für Physik*¹. At face value they provide an explanation for the anomalies that do not invoke Interventionism. On the other hand, they demand that the Universe be far stranger than we had imagined.

If nothing else, Professor Goethe’s analysis provides welcome clarification of the point at issue. Much of the story has now been well rehearsed in our newspapers, albeit not without confusion. We all know how the work of Professor Einstein at Berne, building on that of Professors Planck and Lorentz, initiated a reconsideration of the foundations of physical science. He proposed a breakdown of Newtonian mechanics at both the smallest and the largest scales.

Einstein’s postulate that light is granular² — ‘quantized’, in the phrase now popular — has led to the development of quantum theory by Professors Bohr, Sommerfeld and others. And his proposal that the speed of light in a vacuum is the same in all inertial frames, contrary to canonical Newtonian thought, has led us to a unified view of time and space³, to a new formulation of the equivalence principle, and to a picture of gravity as curved space. So much is clear.

When, three years ago, Professor Goethe first unveiled his alternative theories of both microscopic and cosmological phenomena, based on nothing but classical mechanics, they were received coolly⁴. This is no more than one would have anticipated. Having made the monumental effort to reconstruct physics from the ground up, his colleagues could hardly be expected to take lightly the idea that their efforts had been for nought.

In certain respects, Goethe’s claims held little novelty. Lorentz and Poincaré, after all, had applied nothing more than Maxwell’s equations in a Newtonian framework to explain the failure by Drs Michelson and Morley to detect the luminiferous ether; Goethe’s ‘classical relativity’ added to this in only minor respects. His avoidance of the ultraviolet catastrophe of blackbody radiation, along with his explanation of light-stimulated emission of electrons from metals, was more inventive, invoking inertial limitations to the vibrations of his occluded ether which knitted the two extremes of the theory neatly together. Many believe that this work may now undermine, or at least postpone, Professor Einstein’s candidature for a Nobel award.

How, though, to distinguish Einstein’s Universe from Goethe’s? Discriminating tests were needed. Professor Adrian Perry and his colleague Miss Pearl Dean at Cambridge were the first to oblige. They showed that Professor Eddington’s eclipse observations of last year fitted Einstein’s relativity better than Goethe’s. Order was restored, but fleetingly. Last November, Perry and Dean uncovered archival astronomical observations from the 1839 eclipse that seemed better to match Goethe’s view.

A similar conundrum was shortly thereafter unearthed for the ‘quantum’ phenomena. Einstein’s theory worked best for the new data collected by Perry and Dean; but the notebooks of none other than Professor Weber, Einstein’s former teacher at



JACEY

Zürich, showed that measurements of blackbody radiation made in 1893 supported Goethe, in fine detail, better than they did Einstein himself.

The pattern was repeated wherever one looked: before the turn of the century, the classical theory worked best; after it, relativity and quantum mechanics triumphed. We have for the past several months been wrestling with the uncomfortable notion that the laws of physics themselves shifted around the end of the last century.

It is perhaps a comment on our troubled times that so many were eager to leap to a supernatural conclusion: Interventionism, the idea that some external agency has played a part. Last May, the Vatican issued its now famous statement on the matter. But Professor Goethe has given us what will be to many a more palatable explanation.

He suggests that the entire Universe is capable of undergoing a change of state, akin to that experienced by water as it freezes. Professor van der Waals at Amsterdam showed how these changes can be considered leaps between two energy minima. According to Goethe, the Universe may adopt either a metastable “false vacuum” state in which classical physics holds, or a globally stable “true vacuum” state governed by Einsteinian relativity and quantum theory. We have, he supposes, just witnessed such a transition.

Two aspects of his analysis remain rum, however. The first is that the theory requires the unhappy condition that the Universe is expanding. Professor Einstein has been so kind as to tell me that this may not, in fact, be so hard to credit; but astronomers will surely rebel at the idea.

Second, Goethe’s painstaking analysis of the onset of the Perry–Dean Shift dates it more precisely than previously, and confines it to no more than a day either side of 22 March 1894. Older readers may recall that this was precisely the day on which Professor Kelvin gave his lecture, now oft derided, in which he claimed that all the problems of physics were near to solution. One hopes that even committed Interventionists will not attribute to Professor Kelvin the power to provoke divine retribution for a moment’s hubris. Can we persuade them that “The Lord is subtle, but he is not malicious”?

1. Goethe, *I. Z. Phys.* **22**, 415 (1920).
2. Einstein, *A. Ann. Phys.* **17**, 132 (1905).
3. Einstein, *A. Ann. Phys.* **17**, 891 (1905).
4. Goethe, *I. Ann. Phys.* **29**, 272 (1917).

ATP, pain and a full bladder

Sean P. Cook and Edwin W. McCleskey

Sensations of pain have many causes, from tissue damage to sunburn. Several types of pain are sensed by a mechanism involving ATP — the chemical powerhouse of all cells.

The sensation of pain — whether it originates from a scrape on the knee or a growing tumour — involves a complex interplay of chemical messengers and signalling pathways. Much of this mechanism remains a mystery. But on pages 1011 and 1015 of this issue^{1,2}, two groups describe experiments that mark a milestone in our understanding of how one molecule, the P2X₃ receptor, influences pain. Cockayne and colleagues¹ and Souslova and co-workers² have generated mice in which the P2X₃ receptor — which binds ATP — is knocked out. Their initial analysis of these mice supports the idea that ATP contributes both to the pain resulting from tissue damage and to the discomfort of a full bladder. Unexpectedly, the data also suggest that P2X₃ receptors are involved in sensing warmth.

The P2X₃ receptor is an ion channel — a protein that spans a specific cellular membrane and allows certain types of ions to pass through it. This particular ion channel sits in the outer cell membrane, and is opened when ATP binds to it from the outside of the cell. A connection between the P2X₃ receptor and pain was first uncovered 25 years ago when Bleehen, Hobbiger and Keele³ injected fractions of the cytosol (the contents, excluding the organelles) of red blood cells into human blisters, and found that the fractions that were enriched in ATP caused pain. The authors proposed the general idea that cytosolic ATP — released from damaged cells in damaged tissue — excites pain-sensing neurons (nociceptors).

Over the years, this theory faded into obscurity as other molecular messengers, notably bradykinin and prostaglandins, were implicated in pain signalling. But interest in the 'ATP hypothesis' was revived when researchers cloned the gene encoding the P2X₃ receptor and found that it is expressed highly in nociceptors but is not present outside sensory neurons^{4,5}. This restricted distribution indicated that drugs targeting P2X₃ receptors might inhibit pain without causing side effects. The new work^{1,2} shows both the promise and the peril of this logic.

Cockayne *et al.*¹ and Souslova *et al.*² tested the original ATP hypothesis by damaging tissue in the paws of mice with an injection of a chemical, formalin. Over the next hour, mice lacking the P2X₃ receptors lifted and licked their paws less frequently, indicating that they felt less pain. Injection of ATP itself

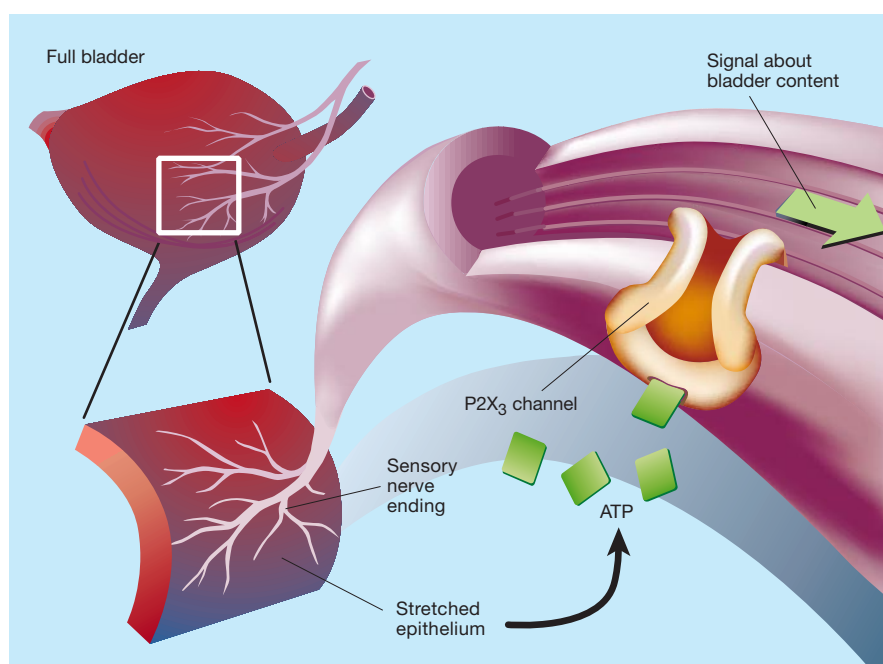


Figure 1 Bladder emptying — why you feel the urge. Cockayne *et al.*¹ and Souslova *et al.*² genetically modified mice so that they lack the P2X₃ receptor, an ion channel that is opened by extracellular ATP and is expressed only on sensory neurons. Cockayne *et al.* found that the mice exhibited greatly diminished responses to the filling and stretching of their urinary bladders — that is, they emptied their bladders less frequently than normal. Previous results⁸ showed that stretching releases ATP from the inside of bladder epithelial cells. Together, these results reveal an unusual form of mechanical sensation: ATP is the molecular messenger released by stretching, and channels made of P2X₃ receptors are the sensors that detect ATP and trigger the neuronal pathway that tells us that the bladder is full and needs to be emptied.

caused pain, which was greatly diminished — but not fully eliminated — in mice lacking P2X₃ receptors. These data support the idea that ATP contributes to the pain caused by tissue damage (although it is not the only such messenger), but also indicate that there might be other nociceptive sensors for extracellular ATP besides P2X₃ receptors. Normal and P2X₃-deficient mice responded similarly to unpleasant compression of the tail or high temperatures applied to the paw — two stimuli that do not cause obvious tissue damage. So, ATP is not a general mediator of all painful stimuli.

Of the other results that came out of these analyses, one of the discoveries made by Cockayne *et al.*¹ is the most convincing, and the most immediately useful: the mice lacking P2X₃ receptors urinated infrequently. Like other hollow organs, the wall of the urinary bladder has a layer of epithelial cells inside a layer of smooth muscle⁶. As the blad-

der fills and stretches, ATP is released from the epithelium^{7,8}. It has been proposed that the released ATP binds to P2X-type receptors on sensory nerve terminals and that ATP is the 'sensory nerve mediator for the degree of distension' of the bladder and other hollow organs⁸ (Fig. 1). Consistent with this theory, when P2X-type receptors are blocked in a urinary-bladder/pelvic-nerve preparation, the nerves fire less frequently in response to bladder distension⁹. The hypothesis now seems to be confirmed by the fact that mice lacking P2X₃ receptors are less sensitive to bladder distension.

Drugs that target P2X₃ receptors should be considered to relieve the pain resulting from some disorders of hollow organs such as the bladder and gastrointestinal tract. But some of Souslova *et al.*'s results² lead us to recommend caution. These authors measured the hypersensitivity that develops in inflamed skin, presumably expecting this

pain to decrease when a putative sensor molecule (the P2X₃ receptor) is absent. Surprisingly, animals lacking P2X₃ receptors became more, not less, hypersensitive after inflammation. Perhaps this result is an artefact of the gene-knockout approach used, and might not occur with short-term use of drugs that block P2X₃ receptors. But this might be wishful thinking, and we should take seriously the possibility that the suppression of these receptors would aggravate inflammatory pain.

Possibly the most unexpected finding from these analyses² is that the P2X₃-receptor-deficient mice no longer showed any neuronal activity in the spinal cord in response to mild skin warming. It is a mystery how we sense pleasant warmth, when the skin temperature rises from 20 °C to 40 °C; no role for P2X₃ receptors has ever been proposed. By contrast, we are fairly certain that the mechanism for sensing harmful heat (45 °C and above) involves the vanilloid receptors^{10,11}. Perhaps P2X₃ receptors are expressed on heat-sensing neurons as well as on pain-sensing ones, and become more active when skin is warmed. However, it would be premature to come to such a provocative conclusion, given that Souslova *et al.* made recordings of neuronal activity in spinal neurons. Spinal neurons receive their input from sensory neurons that extend from the skin to the spine. Sensing the temperature at the skin is just the first in a series of molecular events that carry the signal from skin to spinal cord. A disruption at any point could cause the defects seen in the P2X₃-deficient mice.

These results clearly show the importance of ATP and the P2X₃ receptors in sensing certain types of pain. But many pieces of the puzzle are still missing. Little is known about how ATP is released, for example, or how long it survives after a harmful event. ATP might contribute only temporarily to the pain caused by tissue damage: it is quickly degraded outside cells¹², and channels made of only P2X₃ receptors desensitize in less than a second^{4,5}.

Also, behavioural models for measuring pain are much more sensitive to persistent than to transient signals, so they are better suited to getting to the bottom of the roles of slower molecular messengers, such as bradykinin, rather than ATP. The formalin-induced model of pain used by Cockayne *et al.* and Souslova *et al.* undoubtedly causes tissue damage, yet it is hard to know how it relates to the pain caused by a more natural injury, such as falling off a bicycle. But, despite these limitations, the train of discovery from injecting cytosol into blisters to knocking out P2X₃ receptors from mice has provided unexpected insights into pain transduction, and an unequivocal direction to clinical research.

Sean P. Cook and Edwin W. McCleskey are at the

Vollum Institute, L-474, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA.
e-mails: cooks@ohsu.edu
mccleskey@ohsu.edu

1. Cockayne, D. A. *et al.* *Nature* **407**, 1011–1015 (2000).
2. Souslova, V. *et al.* *Nature* **407**, 1015–1017 (2000).
3. Bleehen, T., Hobbiger, F. & Keele, C. A. *J. Physiol. (Lond.)* **262**, 131–149 (1976).
4. Chen, C. C. *et al.* *Nature* **377**, 428–431 (1995).

5. Lewis, C. *et al.* *Nature* **377**, 432–435 (1995).
6. Dixon, J. S. & Gilpin, C. J. *J. Anat.* **151**, 199–207 (1987).
7. Grygorczyk, R. & Hanrahan, J. W. *Am. J. Physiol.* **272**, C1058–C1066 (1997).
8. Ferguson, D. R., Kennedy, I. & Burton, T. J. *J. Physiol. (Lond.)* **505**, 503–511 (1997).
9. Namasivayam, S., Eardley, I. & Morrison, J. F. B. *Br. J. Urol. Int.* **84**, 854–860 (1999).
10. Caterina, M. J. *et al.* *Science* **288**, 306–313 (2000).
11. Davis, J. B. *et al.* *Nature* **405**, 183–187 (2000).
12. Holton, P. J. *J. Physiol. (Lond.)* **145**, 494–504 (1959).

Astronomy

The red ragged edge

Brian G. Marsden

It is not particularly surprising that the properties of asteroids should change between the inner and outer parts of the main asteroid belt between Mars and Jupiter. After all, this belt marks the transition between the rocky planets of the inner Solar System and the gas giants of the outer Solar System. What is perhaps more unexpected is that a similar transition should occur in the

Kuiper belt—an even greater ring of icy bodies orbiting beyond Neptune. Three years ago, Tegler and Romanishin¹ reported the results of a photometric survey of these distant bodies, showing that they fall into two distinct groups based on their colour, but the reason for the two colours was a mystery. On page 979 of this issue², Tegler and Romanishin present a follow-up survey that demonstrates

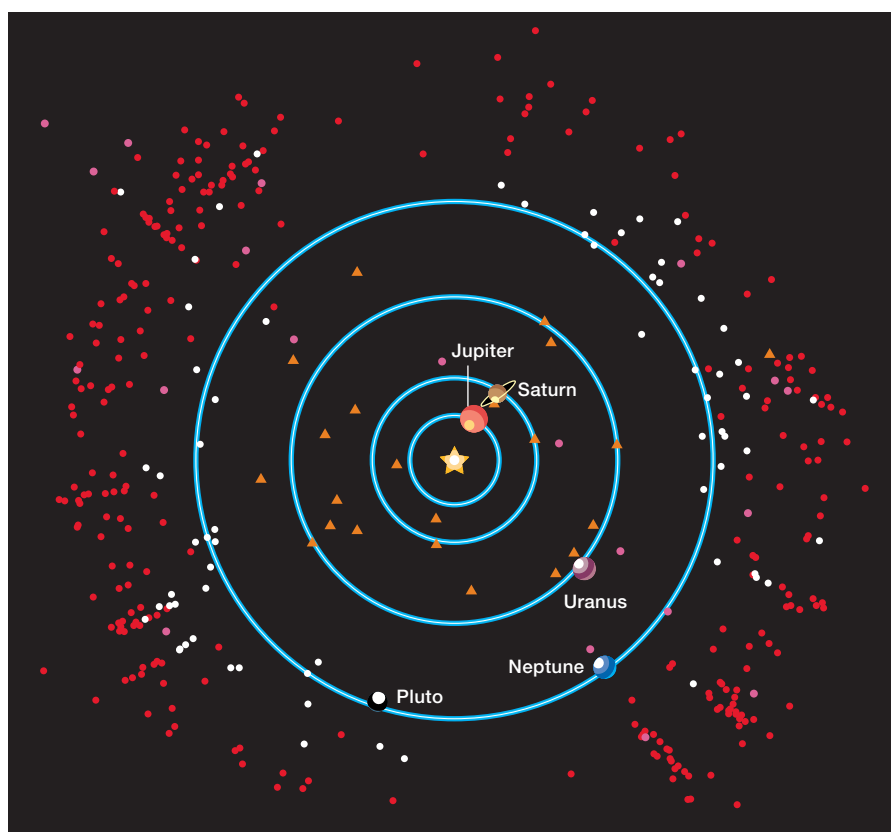


Figure 1 Snapshot of the outer Solar System, as it was on 16 October 2000. The outer Solar System contains the giant planets Jupiter, Saturn, Uranus and Neptune, along with Pluto and other smaller objects. (The inner Solar System, containing the four rocky planets, is too small to be shown.) Pluto's non-circular orbit takes it from its present position at 30 AU from the Sun through the Kuiper belt of icy bodies beyond Neptune to a distance of 50 AU from the Sun. The Kuiper-belt objects (KBOs) include the 'classical' KBOs (red circles), 'Plutinos' (bodies with orbits similar to Pluto; white circles) and 'scattered-disk objects' (magenta circles). Icy comets known as Centaurs (orange triangles) are thought to have escaped from the Kuiper belt, and can be found throughout the outer Solar System. Tegler and Romanishin's photometric surveys^{1,2} of KBOs and Centaurs suggest that the classical KBOs found beyond 40 AU may be 'redder' than the other KBOs.

Neurobiology

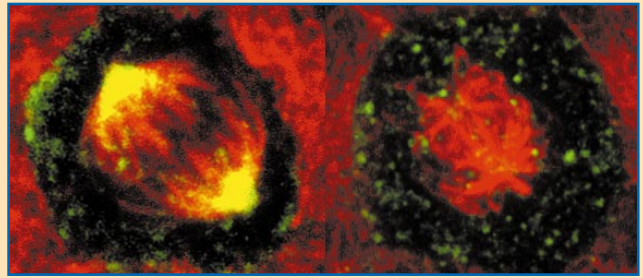
Convolved communications

Normal brains are convoluted in appearance. But those of children suffering from lissencephaly, or 'smooth brain', are not — with tragic consequences (mental retardation, epilepsy and an early death). The cause is mutation of one of the two copies of a gene known as *LIS1*, which means that neurons fail to migrate to the cerebral cortex during development. But what provides the link between *LIS1* mutations and the defects in migration? Three papers in *Nature Cell Biology* tackle the question.

One way of studying the function of a gene is to delete both copies in the embryos of an experimental organism and see what happens. But this can't be done with *LIS1*, because the embryos die. So Zhao Liu *et al.* (*Nature Cell Biol.* **2**, 776–782; 2000)

took another approach. Using the fruitfly *Drosophila melanogaster*, the authors knocked out both *LIS1* genes in developing neurons only in the 'mushroom body', a structure analogous to the human cortex. This resulted in severe abnormalities in the appearance and numbers of neurons (normal adult flies have about 500 neurons, but the mutants had only about 50). Liu *et al.* suggest that the low cell numbers resulted from a defect in cell proliferation.

This interpretation ties in with the results of Nicole Faulkner *et al.* (**2**, 784–791), who found that the division of mammalian cells in culture was dramatically inhibited when *LIS1* was overexpressed or suppressed. Chromosomes did not attach properly to the microtubule-based 'spindle'



that separates duplicated chromosomes, and the orientation of the spindle became random (compare the right-hand spindle in the picture with the normal one on the left). Faulkner *et al.*, and Deanna Smith *et al.* (**2**, 767–775), also showed that *LIS1* protein and components of cytoplasmic dynein — a molecular motor involved in cellular processes that require microtubules — occur

in the same regions of cells.

The authors' studies went beyond those described here. But it seems that *LIS1* and dynein (and dynactin, its companion complex) ensure that developing nerve cells divide at the right time, and in the right orientation. Mutations in *LIS1* apparently disrupt these processes, perhaps leading to a failure of neurons to migrate properly to the cortex. **Amanda Tromans**

a possible link between the colours of these objects and their distances from the Sun.

More than 150 years elapsed between the first calculations of the orbits of asteroids and the first observations that provided a real physical understanding of these bodies. The colour of sunlight reflected off asteroids can be measured through different-coloured filters. Such photometric measurements of several asteroids in the 1950s revealed that their colours fell into two main groups, which were interpreted as differing in surface composition. One group is only slightly redder than the Sun in colour, meaning that their surfaces reflect all wavelengths more or less equally, whereas the other group is significantly redder than the Sun because they absorb more of the blue sunlight. The first group tends to dominate the outer part of the asteroid belt, whereas the redder ones prevail in the inner belt.

During the past decade more than 300 faint objects have been found in the belt beyond Neptune (Fig. 1), which is located at 30 astronomical units from the Sun (1 AU is essentially the Earth–Sun distance). For these objects, only a few years elapsed before knowledge of their orbits was supplemented by physical studies. The idea of a population of small bodies beyond Neptune is variously attributed to Kuiper³ and Edgeworth⁴, although Leonard⁵ wrote of a possible 'ultra-neptunian zone' of bodies within months of the discovery of Pluto. In 1964 Whipple⁶ predicted that Pluto's companions would be dormant comets up to 100 km in radius and between 30 and 50 AU from the Sun. This is how we see the population of 'Kuiper-belt objects' (KBOs) today.

Pluto is now generally considered to be a

member of the Kuiper belt, although its non-circular orbit brings it closer to the Sun than Neptune at its closest approach (its perihelion) and out to 50 AU at its farthest (its aphelion). Fortunately, the 165-year orbital period of Neptune is precisely two-thirds that of Pluto. This '2:3 resonance' restricts the relative configurations of the pair and prevents them from passing within 18 AU of each other⁷. Three of the first six KBO candidates found during 1992–93 have orbits similar to that of Pluto. Known as 'Plutinos', they share the 2:3 resonance that prevents them from getting too close to Neptune⁸. Two others have orbits that are much more nearly circular, with distances ranging between 41 AU at perihelion to 47 AU at aphelion. These objects — sometimes known as 'classical' KBOs — are also safe from encounters with Neptune.

Most of the KBOs found since 1993 follow the same general pattern (Fig. 1). Classical KBOs make up the bulk of the Kuiper belt, being five times more numerous than Plutinos of comparable size. They have mean distances of 41–47 AU, but can reach 38 AU at perihelion and 53 AU at aphelion. There are also objects that avoid Neptune by means of other resonances, and there is a population of 'scattered-disk objects'⁹, which have highly non-circular orbits. These scattered-disk objects bear some relation to the 'Centaur', which are broadly dispersed throughout the outer Solar System (Fig. 1). The first known Centaur, Chiron, has been seen to emit gas and dust like a comet. So it is reasonable to conclude that Centaurs are once-dormant comets that have escaped from the Kuiper belt over many millions of years. In some cases they eventually turn into the short-

period comets found inside Jupiter's orbit.

Tegler and Romanishin's combined surveys^{1,2} include 36 objects, of which 13 are Plutinos, 13 are classical KBOs and 8 are Centaurs. Like the main-belt asteroids, these KBOs and Centaurs appear to come in two colours. One group has surfaces that are best described as 'grey', whereas the other group is noticeably redder. With this larger sample it is possible to look for correlations between colour and distance. The authors find that the Centaurs and Plutinos with perihelion distances of up to 40 AU are evenly split between the two colour groups, but the nine classical KBOs with perihelion distances greater than 40 AU are exclusively 'red'. Four of the classical KBOs have perihelion distances of less than 40 AU, and three of these are grey. But there is a big difference between the angles at which the orbits of these red and grey classical KBOs are inclined to those of the giant planets. Whereas the reds have relatively low orbital inclinations of up to 13°, the three grey ones all have inclinations in excess of 24°, so perhaps they are not so classical after all.

If the high-inclination classical KBOs are a different subpopulation, the fact that all ten 'genuine' classical KBOs are red is a remarkable result, seemingly more exclusive than the colour groupings in the asteroid belt, which later turned out to be more complex. It suggests that the truly classical KBOs are a pristine population that is so safe from encounters with Neptune that it is not subject to significant long-term orbital variations. After all, if these bodies could evolve into Plutinos or Centaurs, some 'reverse' evolution back to classical orbits would be inevitable, in which case it would be hard not

to pollute the population with grey members. In contrast, the grey high-inclination objects, with perihelion distances merging with those of the scattered-disk objects, may have participated in dynamical evolution with the Plutinos and Centaurs.

An alternative theory would allow some dynamical exchange between populations of KBOs, but argue that some physical process ensures that surfaces become redder whenever objects are confined to the ragged edge of the Solar System beyond 40 AU. What this process could be is unclear. An object in the Kuiper belt is so far from the Sun that being at 38 AU rather than 42 AU would appear to make little difference to how much sunlight it receives. Clearly, photometric and orbital

investigations of more KBOs are needed. Nonetheless, in this first attempt at combining the two, KBO studies have come of age. ■

Brian G. Marsden is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: bmarsden@cfa.harvard.edu

1. Tegler, S. C. & Romanishin, W. *Nature* **392**, 49–51 (1998).
2. Tegler, S. C. & Romanishin, W. *Nature* **407**, 979–981 (2000).
3. Kuiper, G. P. in *Astrophysics: A Topical Symposium* (ed. Hynek, J.) 357–427 (McGraw-Hill, New York, 1951).
4. Edgeworth, K. E. *J. Br. Astron. Assoc.* **53**, 181–188 (1943).
5. Leonard, F. C. *Leaflet Astron. Soc. Pacific* **30** (1930).
6. Whipple, F. L. *Proc. Natl Acad. Sci. USA* **51**, 711–718 (1964).
7. Cohen, C. J. & Hubbard, E. C. *Astron. J.* **70**, 10–13 (1965).
8. Marsden, B. G. in *Completing the Inventory of the Solar System* (eds Rettig, T. W. & Hahn, J. M.) 193–207 (Astron. Soc. Pacific, San Francisco, 1996).
9. Luu, J. *et al.* *Nature* **387**, 573–575 (1997).

Behavioural ecology

Why are some males dull?

Tore Slagsvold

Birds vary extensively in plumage colour, both between and within species. The males of a species are usually brighter than the females, and older males tend to be brighter than younger birds. But males of the same age within a single population may also show great variety in colour. In such cases, the more brightly coloured males are usually socially dominant and are the mating partners of choice for females¹. Why, then, hasn't natural selection eliminated the duller varieties?

On page 1000 of this issue, Greene and colleagues² suggest one explanation to the puzzle. They studied the lazuli bunting (*Passerina amoena*), a beautiful small finch of North America. Adult lazuli bunting males have turquoise-blue upper parts with a pale cinnamon colour across the breast and sides. Some yearling males are also brightly coloured, but many are dull brown and female in appearance, while still others have intermediate brightness (see the photographs on page 1000). Greene *et al.* show that both the brightly coloured and the dull yearling males sire more chicks than yearlings of intermediate colours. It seems that nature favours the extremes at the expense of those in between. This phenomenon is called disruptive selection, and has been seen only a few times in animals.

Why did the brightly coloured and dull birds succeed where those of intermediate colouring failed? Greene *et al.* found that both bright and dull yearling males occupied territories that included suitable shrub cover, whereas the intermediate yearlings settled on sparsely vegetated areas, where hardly any female would breed. Brightly coloured yearling males succeeded because they possessed the fighting ability necessary

to occupy a good territory in the face of competition from adult males. In contrast, dull yearlings obtained a good territory because adult males tolerated their presence. Yearlings of intermediate colouring were too bright to be tolerated by adult males, but were not aggressive enough to fight their way to a good site.

The next question, then, is why adult males tolerated dull yearling males. One possibility is that those adults in possession of good territory had problems in recognizing the true sex of prospecting dull males. But no support for this 'female mimicry' hypothesis³ has yet been found in the lazuli bunting⁴. Instead, the authors suggest that dull males were accepted because they represented no threat to the mating success of adult males, or to the number of chicks that adult males sired. The dull males might also act as a buffer against even brighter males, which females tend to prefer.

Because of this bias in female preference, adult males may even sire extra young by copulating with the mates of their drab male neighbours. Indeed, DNA 'fingerprinting' showed that the number of chicks sired by adult males was positively correlated with the proportion of immediate neighbours that were dull yearlings. So, both parties seem to benefit from this peculiar system — adult males by 'paternity insurance' and by siring extra young, and dull males by being allowed to occupy a high-quality territory and thereby obtain a mate. In addition, previous occupancy of a good territory helped drably coloured owners to obtain a high-quality territory in the subsequent year. The study provides rare evidence for cooperation between males — a territorial arrangement that



100 YEARS AGO

In the *Atti dei Lincei*, ix. 5... Prof. Grassi describes experiments carried out by a committee with the assistance both of the Italian Government and of the Mediterranean Railway Company, with a view to the prevention and cure of malaria in infected districts. The experiments were carried out in the plains about Paestum, which have long been known as a hotbed of malaria ('malaricissima' is the epithet Grassi applies to the region), and fell into two categories, namely, cure of the disease by the use of quinine, and protection from the bites of *Anopheles claviger* by the use of wire gauze as a covering for windows, doors and even chimneys of houses, the inhabitants of which were required to remain indoors from before sunset till after sunrise, or to go about covered with veils at night. By thus preventing mosquito bites, it was found that the malarial regions could be safely inhabited even at the season when the fever was at its height, and under such conditions the district might be made as healthy as any part of Italy. From *Nature* 25 October 1900.

50 YEARS AGO

An article in the *South African Archaeological Bulletin* (5, No. 18; June 1950) makes sad reading. At Saulspoort, in the Bethlehem district, rock-shelter paintings occur and an important 'gisement' was identified... The work of excavation was begun; but during a period of the workers' absence intruders arrived who just 'hogged' the site and left a yawning pit where the section should have been. The problem of controls is a difficult one. Too rigid rules defeat their own object. It is, in a subtle way, the general interest in archaeology among all classes that makes the subject live and enables the few professionals to continue the study. Without this general interest the subject would in practice wither. A certain freedom to explore must therefore be given to the amateur; but, of course, stories such as this one are major tragedies. Perhaps the problem could best be tackled through the schools. If the young folk were taught to realize that Stone Age sites are not innumerable and should be respected, that 'hogging' sites is a crime and that excavations should only be attempted either under a competent excavator or after having already had experience in digging, then such grievous happenings as occurred at Saulspoort would no longer occur. From *Nature* 28 October 1950.

increases the reproductive success of both parties.

Strong disruptive selection, such as that seen here, sometimes leads to individuals with alternative tactics having equal reproductive success. A good example is that of coho salmon (*Oncorhynchus kisutch*), in which some small males persist in the population because they sneak in to fertilize eggs when females spawn on the territories of large males⁵. By contrast, the duller yearling male lazuli buntings sired fewer young than the brightest yearlings. From this finding, and from the fact that most yearling males had an intermediate and apparently suboptimal plumage colour, the authors suggest that genetics plays a small part in male coloration. They propose that environmental constraints — such as the condition of the bird during moulting — are important in determining coloration.

The study by Greene *et al.*² was a success thanks to their integration of behavioural, ecological and genetic analyses, which enabled them to link variation in plumage colour with variation in an ecological factor (territory quality) and with complex patterns of mating success. This approach should be followed in the future. Another question for future studies is whether an association with bright males in a socially monogamous species — such as the lazuli bunting — increases the mating success of the dull males in the way predicted by the so-called hotshot model⁶. In this model, bright males (hotshots) often attract several females, some of which might settle with the dull males.

Greene *et al.*'s work offers an explanation for intrasexual variety in male sexual traits in one species. But their explanation may not be universally applicable. In the lazuli bunting there is a high level of 'extra-pair' paternity — dull yearling males sire fewer young by their mates as a result of competition from brighter males. But this particular pattern of cuckoldry does not occur in all species that exhibit variation in male sexual traits^{7,8}. Greene *et al.* suggest that the heritability of coloration is low in the lazuli bunting, a phenomenon that has been found in some other birds but not in others¹. Explaining such differences between species will be a great challenge. One thing we need to know more about is how genes interact and are expressed in different genetic and ecological environments. For instance, do genes that enhance sexual attractiveness have positive or negative effects on survival^{9,10}? And are there any mutations that are beneficial to one sex but harmful to the other¹¹?

Tore Slagsvold is in the Department of Biology, University of Oslo, PO Box 1050 Blindern, N-0156 Oslo, Norway.
e-mail: tore.slagsvold@bio.uio.no

1. Andersson, M. *Sexual Selection* (Princeton Univ. Press, 1994).

2. Greene, E. *et al. Nature* **407**, 1000–1003 (2000).

3. Rohwer, S., Fretwell, S. D. & Niles, D. M. *Am. Nat.* **115**, 400–437 (1980).

4. Muehler, V. R., Greene, E. & Ratcliffe, L. *Behav. Ecol. Sociobiol.* **41**, 281–290 (1997).

5. Gross, M. R. *Nature* **313**, 47–48 (1985).

6. Beehler, B. M. & Foster, M. S. *Am. Nat.* **131**, 203–219 (1988).

7. Hill, G. H., Montgomerie, R., Roeder, C. & Boag, P. *Behav. Ecol. Sociobiol.* **35**, 193–199 (1994).

8. Lifjeld, J. T., Slagsvold, T., Dale, S. & Ellegren, H. *Auk* **114**, 112–115 (1997).

9. Petrie, M. *Nature* **371**, 598–599 (1994).

10. Brooks, R. *Nature* **406**, 67–70 (2000).

11. Rice, W. R. *Nature* **381**, 232–234 (1996).

El Niño

Clues from corals

Robert B. Dunbar

Weather forecasters now enjoy increasingly accurate predictions of climate events linked to the El Niño/Southern Oscillation system (ENSO), ranging from powerful hurricanes in the Atlantic to devastating droughts in southern Africa and Indonesia. Lead times of up to a year can make a big difference in assessing and preparing for economic and agricultural risks. Yet surprises still come with every El Niño. The magnitudes of the two largest events of the twentieth century, in 1982–83 and 1997–98, were not readily forecast (or hindcast for that matter), nor was the extended mild El Niño of 1990–94. One ingredient still missing from the recipe for prediction is a knowledge of the nature and consequences of decadal and longer climate variability.

On page 989 of this issue, Urban *et al.*¹ provide an intriguing view of tropical variability from an unexpected source — a coral record spanning the past 155 years. The authors find that the timescale of the ENSO cycle is not only highly irregular but also that its irregularity may be linked to subtle changes in the mean background or time-averaged state of the ocean.

In the 34 years since Jacob Bjerknes first proposed a mechanistic link between ocean temperatures in the equatorial Pacific and much larger-scale patterns of atmospheric circulation², one of the great triumphs of research into climate dynamics has been the formulation of predictive models of the ENSO system. The basic physics of ENSO became clear after the introduction in the 1980s of a costly but effective monitoring

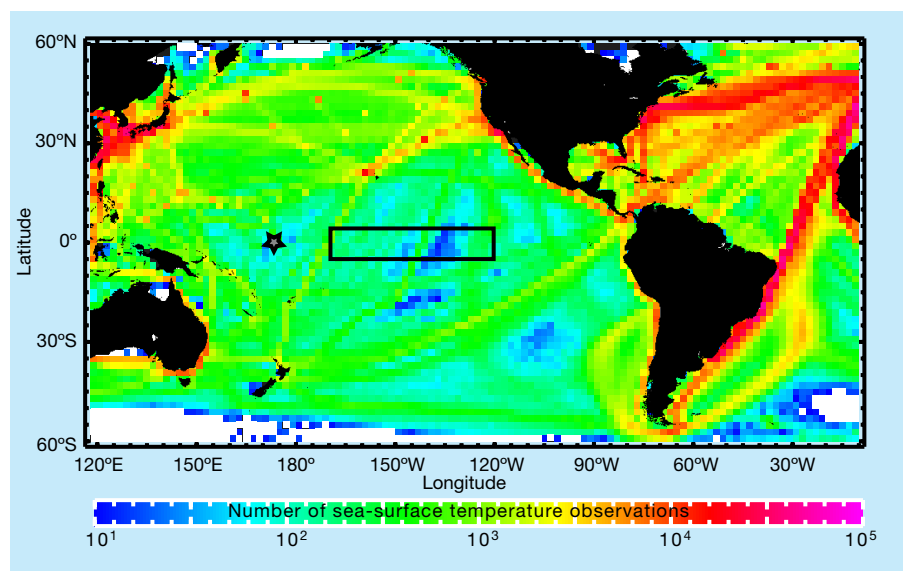


Figure 1 Taking the temperature of the oceans. Number of ship-based observations of sea surface temperature taken between 1900 and 1950 for each 2° latitude by 2° longitude cell of the global ocean. These observations form the core of our instrumental record of surface-ocean variability. Surface water temperatures at the Maiana Atoll (star) and the NINO 3.4 region (rectangle) are particularly sensitive to the state of the El Niño cycle, and temperatures at NINO 3.4 are increasingly used in forecasting. Instrumental records are most common along early twentieth-century shipping lanes and throughout most of the North Atlantic. Within most of the equatorial Pacific there are far fewer measurements over the same period. In any case, highly reliable records of tropical ocean variability extend back only 30 to 50 years. Studies of longer-term interannual to decadal climate variability will rely increasingly on biological archives, such as the coral skeletons from the Maiana Atoll described by Urban *et al.*¹.

system in the tropical Pacific. The ocean and atmosphere in the tropics are tightly coupled — meaning that they are constantly exchanging heat and water, and therefore respond to each other over timescales of weeks to months. Nowhere is this coupling more dynamic than in the Pacific, where the warmest ocean waters on Earth are found west of the dateline. The eastern Pacific is normally cool, and the resulting temperature gradient helps to set up strong trade winds that cause ocean currents near the Equator to flow mostly westward, all the time accumulating heat and so reinforcing the westward rise in temperatures.

This pattern, although normal, is not stable. El Niños begin when the warm water in the west extends more eastward than usual or when the trade winds in the western Pacific weaken or even reverse. Relatively minor events such as these can lead to a feedback between the ocean and atmosphere whereby weaker winds allow warm water to penetrate ever further eastward, weakening the trade winds even more.

During El Niños, warm water moves eastward as a series of waves, not like the small wind-driven waves that ride the surface of the ocean but as very long and slowly moving features that extend deep into the upper ocean. During neutral periods between El Niño and its counterpart La Niña (when the eastern Pacific is unusually cool and the trade winds are strong), the Pacific is always poised on the brink of change and is easily influenced by noise in the climate system — this is the least predictable time of the ENSO cycle. But once a change begins, feedback between the ocean and the atmosphere orchestrates a predictable progression of events lasting up to 18 months. The entire cycle repeats every four years or so. Models of ENSO simulate the air–sea interactions and the movements of heat within the ocean by way of currents and large equatorial waves.

One factor missing from these models is decadal-scale climate variability, the physics and cause of which are still unknown. For example, during 1976–77 much of the northern and tropical Pacific warmed abruptly and stayed warm for the next two decades, with large impacts on west-coast fisheries and long-term average rainfall over parts of North America³. As the ENSO cycle involves large redistributions of heat in the tropical Pacific, is it just coincidence that unusually strong or extended El Niños have occurred since 1977? If the world continues to warm, might we expect future El Niños to be even stronger? Or how might the ENSO cycle respond if decadal variability brings cooler conditions to the eastern Pacific, even against a backdrop of global warming?

The answers to these questions remain elusive because of scarce instrumental observations before the 1950s — we simply don't know much about the state of the tropical



Figure 2 Drilling for answers. Scuba divers use an underwater hydraulic drilling system to recover a core from a reef coral (*Porites lobata*) located at Clipperton Atoll in the eastern tropical Pacific. Massive reef corals grow at rates of about 1 centimetre per year. As they grow, the carbonate skeleton retains chemical and isotopic signatures of ocean temperature and salinity, sometimes spanning several centuries. Urban *et al.*¹ now describe a coral record that documents large changes in the El Niño/Southern Oscillation climate system over the past 155 years.

ocean during most historical El Niño and La Niña events (Fig. 1). But biological archives of tropical variability, such as tree rings and corals, are now extending the instrumental record far into the past (Fig. 2).

Urban *et al.*¹ use oxygen-isotope ratios in the carbonate skeletons of reef corals from Maiana Atoll in the central Pacific to reconstruct past changes in sea surface temperature and salinity. This site is highly sensitive to the ENSO cycle and during the late twentieth century the Maiana coral record is nearly indistinguishable from instrumental records. But whereas the number of instrumental observations diminishes before 1950, the coral record extends back to 1840. The Maiana record suggests that warming and possible freshening (caused by greater rainfall) of the central tropical Pacific over the past 150 years was not steady and incremental, but occurred mostly in two steps — a small shift in the early twentieth century, and a larger shift around 1976–77. The four-

year El Niño cycle we see today is a recent feature extending back only to the 1950s. In the cooler and drier tropical ocean of the late nineteenth century, ENSO-like cycles lasted about 10–15 years. These were replaced by strong three-year cycles that coincided with the warming step in the early twentieth century, followed by three decades of weak interannual variability until 1950.

These new coral results show that decadal variability in ENSO is the 'normal' situation and that the length of the El Niño cycle varies with even small changes in tropical temperatures, albeit in a complex fashion. This is a challenge for numerical simulations of ENSO, which were developed using observations from only the most recent ENSO regimes. It also highlights the need for a better understanding of the causes and mechanisms of decadal climate change. The Maiana coral and other recent studies^{3,4} suggest that the 1976–77 climate shift is the most prominent of several climatic steps during the past century. If these steps are part of a poorly known climate oscillation, they may one day be predictable. Of course, it is also likely that ENSO will feed back in some way onto the decadal variability.

One of the chief lessons arising from 20 years of research into the variability of the atmosphere and ocean is that climate phenomena are highly interconnected across a large range of timescales and latitudes. Indeed, it is this interconnectedness that makes climate prediction so difficult. The Maiana record reveals that decadal variability was most dominant in the late 1800s. This period should be a target for future studies because the clear decadal signal may be more easily linked to possible causative mechanisms.

What does the coral record tell us about ENSO and global warming? Although it supports the notion that ENSO timescales respond to changes in background ocean temperatures, the simple fact that ENSO cycles as seen in the coral record varied from three to more than ten years before the 1920s argues against a simple human influence on modern El Niño variability. Perhaps more important is the observation that the warming and freshening of the central Pacific since the late 1970s is unique over the entire 155-year record. If we include other coral records from the Indian and Pacific oceans^{5–8}, then the recent warm and wet conditions appear unique over the past 200 to 300 years. It seems likely that ENSO will respond to further global warming⁹. Our ability to predict the nature of this response will rely increasingly on biological archives that hold the secrets of decadal climate change. ■

Robert B. Dunbar is in the Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305-2115, USA. e-mail: dunbar@stanford.edu

1. Urban, F. E., Cole, J. E. & Overpeck, J. T. *Nature* **407**, 989–993 (2000).

2. Bjerknes, J. *Tellus* **8**, 820–829 (1966).
3. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
4. Chao, Y., Ghil, M. & McWilliams, J. C. *Geophys. Res. Lett.* **27**, 2261–2264 (2000).
5. Gagan, M. K. *et al. Quat. Sci. Rev.* **19**, 167–182 (2000).
6. Cole, J. E., Dunbar, R. B., McClanahan, T. R. & Muthiga, N. A. *Science* **287**, 617–619 (2000).
7. Linsley, B., Dunbar, R. B., Wellington, G. M. & Mucciarone, D. A. *J. Geophys. Res.* **99**, 9977–9994 (1994).
8. Quinn, T. M. *et al. Paleoceanography* **13**, 412–426 (1998).
9. Federov, A. V. & Philander, S. G. *Science* **288**, 1997–2002 (2000).

Photosynthesis

Carbon fix for a diatom

Ulf Riebesell

Water — the elixir of life — is often a scarce resource. Many land plants in warm and arid climates conserve water by temporarily closing their 'breathing tubes', the so-called stomata. This strategy has a drawback, however, in that it reduces the flow of CO₂ from the atmosphere into the plant, so lowering its photosynthetic capacity. To solve the dilemma, these plants have devised a mechanism to optimize CO₂ uptake. The key component is a CO₂ storage and transport compound that contains four carbon atoms and gives this process its name — the C₄ pathway. On page 996 of this issue¹, Reinfelder and colleagues show for the first time that a marine microalga, the coastal diatom *Thalassiosira weissflogii*, uses the same C₄ pathway as some of its terrestrial counterparts.

But why would a plant suspended in the ocean use a mechanism otherwise employed to conserve water? The answer is simple: just

like a terrestrial plant holding its breath, an alga growing in the sea can suffer from a shortage of CO₂. Although inorganic carbon is abundant in the sea, 90% of it is in the form of bicarbonate ions (HCO₃[−]). Less than 1% is present as CO₂, which is the form required by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), the enzyme primarily responsible for photosynthetic carbon fixation.

At typical concentrations of CO₂ in sea water, Rubisco operates far below its optimum. Algae usually overcome the problem by increasing the CO₂ concentration at the site of carbon fixation through the active uptake of CO₂ or HCO₃[−], rather than relying on diffusion². Carbonic anhydrase, a zinc-containing enzyme that catalyses the conversion between CO₂ and HCO₃[−], is a central component of this so-called carbon-concentrating mechanism³. Reinfelder *et al.*¹ find the strongest expression of C₄

carbon fixation after applying stress to the carbon-concentrating mechanism by reducing CO₂ availability, lowering zinc concentrations (which is thought to reduce carbonic anhydrase activity) or directly inhibiting carbonic anhydrase. The implication is that C₄ fixation operates to feed Rubisco when the carbon-concentrating mechanism cannot.

The Rubisco carbon fixation and C₄ pathway are competitive processes, so it is essential that they are segregated in either space or time. Terrestrial C₄ plants such as sugar cane or maize have chosen the first solution. They fix CO₂ in a four-carbon molecule, malate, in one type of cell, then release it for fixation by Rubisco in another cell type (Fig. 1a, overleaf). Many succulent plants have opted for the temporal solution: their stomata open only at night, when water loss due to transpiration is lowest. CO₂ is bound in malate in the dark and is released for fixation by Rubisco in the day (Fig. 1a). One of the novelties of Reinfelder and colleagues' results is that both processes occur simultaneously in a single-celled organism. According to their data, C₄ carbon fixation may be confined to the cytoplasm, spatially segregated from the Rubisco process, which occurs in the chloroplasts (Fig. 1b).

Like other phytoplankton groups, diatoms can usually satisfy their carbon requirement entirely by way of the 'classical' carbon-concentrating mechanism^{2,4}. So the findings of Reinfelder *et al.*¹ raise the intriguing question of why diatoms have

Nanotechnology

Crossroads in carbon

The smallest electronic device could be based on just one single molecule. Carbon nanotubes — flexible, hollow nanowires with versatile electronic properties — have already proven themselves as miniature diodes and transistors. In a paper in *Applied Physics Letters* (**77**, 2530–2532; 2000), C. Rao and colleagues from the Jawaharlal Nehru Centre for Advanced Scientific Research in India now demonstrate an efficient method for synthesizing a more advanced structure from carbon nanotubes: Y-junctions. Such structures could be used in new types of molecular devices.

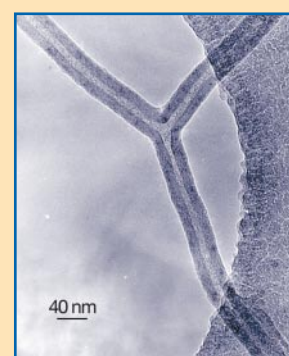
Carbon nanotubes are known for their remarkable property to be either a semiconductor or a metal, depending on their diameter and the winding of the carbon sheet from which the nanowire is made. A sharp bend in a nanotube can actually be

thought of as a junction between two nanotubes with different electronic behaviour and so provide a transition from semiconductor to metal over a distance of just a few nanometres. Such sharply bent tubes have already been used as molecular diodes. But scientists are always on the lookout for more complex structures based on carbon nanotubes. A Y-junction can be thought of as a connection between three different carbon nanotubes, which could form, for example, a microscopic metal–semiconductor–metal contact.

In previous attempts to construct complex junctions, two nanotubes have been crossed or Y-shaped templates have been used to laboriously mould a junction from a single nanotube. A simple method to produce carbon nanotubes is pyrolysis of organic molecules. In

this process, carbon-containing molecules are decomposed at high temperatures, using appropriate catalysts. Rao and co-workers have finely tuned this method to create their Y-tubes, with a 70% yield. They decompose nickelocene, an organic molecule containing a nickel atom, along with another organic molecule, thiophene, at a temperature of 1,273 K. An electron microscope image of the product (shown here) reveals that the Y-shaped nanotubes are multi-walled and have an outer diameter of about 40 nanometres. The angle between the upper arms is almost 90°.

One of the current objectives in nanotube synthesis is to have control over the electronic properties of the end product. Although the electronic structure of these Y-junctions is not known exactly, initial measurements by Rao and co-



workers show that their Y-junctions can behave like diodes. This work is still preliminary, but it will inspire further studies into making three-point nanotube junctions with specific semiconductor–metal transitions. Such molecular junctions will be useful building blocks in the continuing miniaturization of complex electronic devices.

Liesbeth Venema

two means (the C_4 pathway and the carbon-concentrating mechanism) to the same end. But under what particular circumstances might C_4 carbon fixation cut in? Here's some speculation.

The C_4 pathway is of greatest advantage in plants requiring temporary storage of CO_2 . This is obviously the case for plants in which Rubisco carbon fixation occurs partly or entirely when the stomata are closed. In aquatic environments, on the other hand, CO_2 and HCO_3^- availability is comparatively stable over time. Here, it is the photosynthetic fixation of carbon that is highly variable because of the unpredictable amount of light available to phytoplankton. Superimposed on the diurnal solar cycle, phytoplankton experience large fluctuations in light intensity, because they may be swept up and down in the water column by vertical mixing within the ocean surface layer. Temporary storage of CO_2 at times of low irradiance, and its release at times of increased photosynthetic carbon demand during high illumination close to the surface, could be used for intermittent supplementation of carbon from the carbon-concentrating mechanism. Interestingly, diatoms typically dominate the phytoplankton in turbulent

waters with relatively deep mixing of the water column⁵. Using two processes, one to ensure a steady flow of CO_2 (the carbon-concentrating mechanism), the other to provide an extra dose during periods of high CO_2 use (C_4), may be part of a strategy that allows diatoms to flourish in a fluctuating environment.

From an evolutionary point of view, diatoms are the most likely phytoplankton group to have developed a C_4 pathway. As Reinfelder *et al.*¹ point out, diatoms underwent their main evolutionary diversification during an era when atmospheric CO_2 concentrations were relatively low. In contrast, most other groups of phytoplankton such as cyanobacteria, green algae and dinoflagellates evolved earlier⁶, when there were higher levels of CO_2 . Increased selective pressure for more efficient use of CO_2 may also be reflected in the higher CO_2 specificity of diatom Rubisco compared to that of more ancient groups⁷. If C_4 photosynthesis occurs in marine diatoms generally, its evolution predated that in terrestrial C_4 plants by a long time. So unravelling the evolutionary roots of C_4 photosynthesis in algae opens an exciting line of research.

To assess the implications of marine

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

C_4 photosynthesis, we must question the importance of this process in the natural environment. Reinfelder *et al.* grew *T. weissflogii* under zinc concentrations that were far lower than would be experienced by coastal diatoms⁸. This may have placed unusual stress on the carbon-concentrating mechanism, thus inducing C_4 metabolism in their experiments.

If C_4 photosynthesis does account for a significant portion of marine carbon fixation, it will affect various aspects of marine ecology and biogeochemistry. For example, the presence of the C_4 pathway is likely to influence algal sensitivity to changes in CO_2 concentrations. As in terrestrial ecosystems, C_4 photosynthesis may therefore be a factor that is shaping species distribution and succession if it occurs in only some members of the phytoplankton. It could operate both on geological timescales and in response to the present rise in atmospheric CO_2 concentrations.

Furthermore, C_4 photosynthesis lowers the 'discrimination' against the heavy isotope ^{13}C during photosynthetic carbon fixation, so the organic matter produced through the C_4 pathway is comparatively less depleted in ^{13}C . If this pathway turns out to be quantitatively significant in the production of marine organic matter, it will necessitate a re-interpretation of carbon isotope ratios in both phytoplankton and ocean sediments. The latter are used as proxy data in various applications. Examples are in studying food-web structure and the marine carbon cycle, and in attempting to reconstruct CO_2 levels or phytoplankton growth rates in the past.

So much may (or may not) stem from the results of Reinfelder *et al.* At the very least, however, the existence of C_4 photosynthesis in diatoms demonstrates their tremendous flexibility in responding to environmental variability.

Ulf Riebesell is at the Alfred Wegener Institute for Polar and Marine Research, PO Box 120161, 27515 Bremerhaven, Germany.

e-mail: uriebesell@awi-bremerhaven.de

1. Reinfelder, J. R., Kraepiel, A. M. L. & Morel, F. M. M. *Nature* **407**, 996–999 (2000).
2. Badger M. R. *et al.* *Can. J. Bot.* **76**, 1052–1071 (1998).
3. Sültmeyer, D. *Can. J. Bot.* **76**, 962–972 (1998).
4. Raven, J. A. *Adv. Bot. Res.* **27**, 85–209 (1997).
5. Harris, G. P. *Phytoplankton Ecology: Structure, Function, and Fluctuation* (Chapman & Hall, London, 1986).
6. Van den Hoek, C., Mann, D. G. & Jahns, H. M. *Algae: An Introduction to Phycology* (Cambridge Univ. Press, 1995).
7. Tortell, P. D. *Limnol. Oceanogr.* **45**, 744–750 (2000).
8. Bruland, K. W. *Limnol. Oceanogr.* **34**, 269–285 (1989).

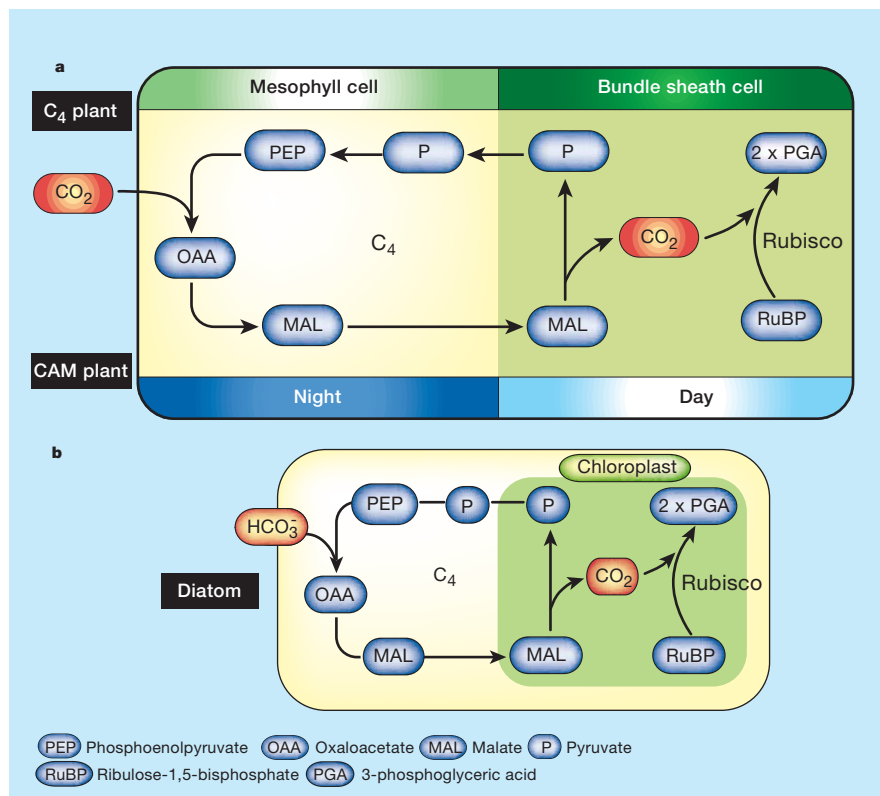


Figure 1 Fixation of CO_2 in the C_4 pathway. a, In land plants the competitive processes of C_4 (left) and Rubisco (right) carbon fixation are segregated. Segregation can be spatial, as in sugar cane and maize; or it can be temporal, as in the reaction cycle, known as crassulacean acid metabolism (CAM), that occurs in many succulent plants. b, Equivalent events in the single-celled diatom *Thalassiosira weissflogii*, as proposed by Reinfelder *et al.*¹. Here the two processes can happen simultaneously because one occurs in the cytoplasm and the other in chloroplasts. After further reactions, the product of the biochemical cycles shown here, PGA, can be used to produce fatty acids, amino acids or sugars.

Insecticides and mosquito-borne disease

Insecticide resistance in mosquitoes can also interfere with developing parasites.

The primary means of controlling mosquito-borne diseases such as malaria and filariasis is still by residual spraying with insecticides. Here we show that insecticide-resistant *Culex quinquefasciatus* mosquitoes are less likely to transmit filariasis than their insecticide-susceptible counterparts. If this surprising finding extends to other combinations of insect species, insecticide-resistance mechanisms and disease, it could have widespread consequences for the control of vector-borne disease.

The development of insecticide resistance in mosquito vectors is common¹. Insecticide resistance is assumed to increase the likelihood of disease transmission by mosquitoes by increasing the population size and allowing mosquitoes to live longer in the presence of insecticide. We have tested the validity of this assumption in *C. quinquefasciatus* from Sri Lanka.

Culex quinquefasciatus uses one predominant resistance mechanism that occurs in more than 80% of insecticide-resistant *Culex* worldwide^{2,3} and which originated in one population before spreading rapidly³. This resistance depends on the stable germline amplification of two esterase enzymes and an aldehyde oxidase encoded on 30 kilobases of DNA⁴; up to 80 copies of this DNA amplicon can be present per cell⁵.

One of the esterases found in resistant insects is expressed at very high levels in the mosquito gut, subcuticular layer, Malpighian tubules and salivary glands, resulting in a change in the redox potential in these cells⁶. As most mosquito-borne parasites must pass through these tissues to complete their development, it is possible that parasite survival, and hence the vectorial capacity of the insect, may be directly affected by the insecticide-resistance status of the insects.

Lymphatic filariasis, caused by the parasitic worm *Wuchereria bancrofti*, is an endemic disease in Sri Lanka⁷. Since 1974 mosquitoes have been controlled by spraying their breeding sites fortnightly with the organophosphate fenthion⁸. There is a moderate frequency of esterase-mediated resistance in *C. quinquefasciatus* throughout the island, although this mechanism is not particularly effective against fenthion⁸.

We collected blood-fed female *C. quinquefasciatus* from the houses of filariasis patients in seven districts. Mosquitoes were snap-frozen within 48 hours of taking their blood meal and analysed for parasite load by quantitative polymerase chain reaction (PCR) and for insecticide resistance by biochemical assay using *p*-nitrophenyl-

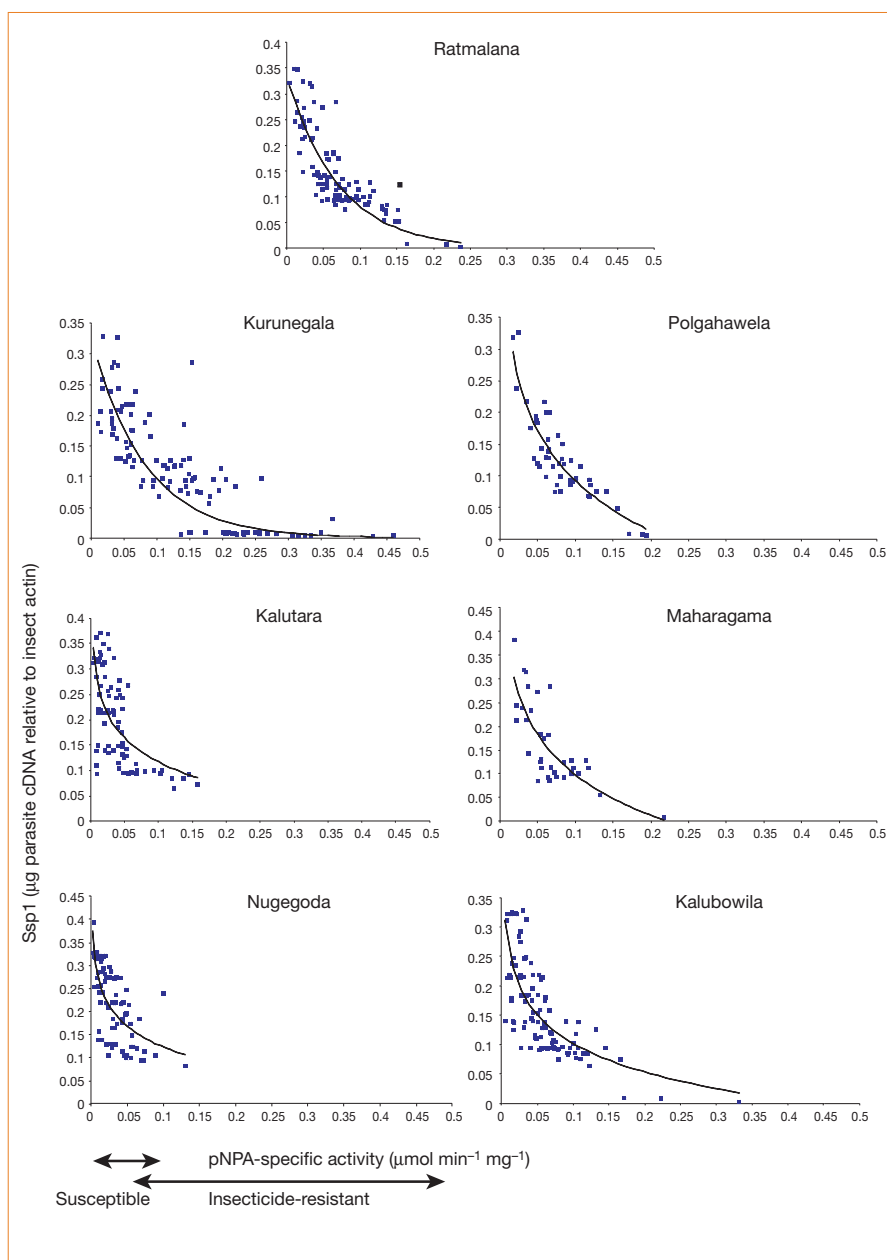


Figure 1 Relation between insecticide-resistance status, as determined by esterase activity using *p*-nitrophenylacetate (pNPA), and quantity of *Wuchereria bancrofti* RNA (as determined by the quantity of parasite Ssp1-repeat complementary DNA sequence relative to the insect actin cDNA in the same extraction) carried by individual field-collected mosquitoes less than 48 h after an infected blood meal; insects were collected from seven locations in Sri Lanka (for further details, see Supplementary Information). Arrows indicate ranges of esterase activity in insecticide-susceptible and in insecticide-resistant insects.

acetate. For quantitative PCR, we used primers specific to *W. bancrofti*⁹, standardized against a mosquito actin control¹⁰ from a complementary DNA template produced from RNA extracted from whole infected mosquitoes.

Almost 80% of the insecticide-susceptible and resistant mosquitoes that we collected were infected with *W. bancrofti*. However, at all seven localities there was a

strong negative correlation between esterase activity, as determined with *p*-nitrophenylacetate, and parasite RNA levels (Fig. 1). The reduction in parasite RNA in insecticide-resistant mosquitoes was not due to differential mortality of parasite-infected resistant insects, as insecticide-resistance gene frequencies were similar in infected mosquitoes, in field-caught mosquito larvae, and in uninfected mosquito adults;

blood-meal sizes were also comparable for the resistant phenotype.

Artificially feeding insecticide-resistant and insecticide-susceptible mosquito colonies (PelRR and PelSS, respectively¹¹) with blood infected by *W. bancrofti* to an intermediate level of parasitaemia — which should result in the infection of mosquitoes without substantial insect mortality — produced stage-L3 infective parasite larvae after 12 days in 76% of PelSS females ($n=250$), but no larvae in any of the PelRR females ($n=200$). Our results indicate that an increase in esterase activity could affect the development of stage-L1 *W. bancrofti* larvae, which may be arrested in the gut cells of insecticide-resistant but not insecticide-susceptible mosquitoes.

Filarial infection severely damages the mosquito host, often killing it. The spread of esterase-based insecticide resistance in field populations of *C. quinquefasciatus* may therefore be influenced by selection pressures for both insecticide detoxification and reduction of the microfilarial burden. Similar esterase-based insecticide-resistance mechanisms have been selected in field populations of the malaria vectors *Anopheles albimanus*¹² and *A. culicifacies*¹³, which could directly affect the transmission of malaria.

Evolution

Paying for sex is not easy

Explaining the maintenance of sexual reproduction remains one of the greatest challenges for biology, with more than 20 hypotheses having been advanced so far¹. Doncaster *et al.*² have proposed another possible explanation, but we question the novelty and importance of their suggested mechanism.

First, does this model² provide a new mechanism to help explain sex? The model assumes that different genotypes exploit different parts of the environment, and that an asexual clone is not able to occupy all the environmental niches that are open to a sexual population. This provides an advantage to sex, and allows coexistence between sexuals and asexuals.

However, the assumptions are the same as those of the 'tangled bank' mechanism for the maintenance of sex^{1,3–5}. The message that has emerged from previous considerations of this mechanism is therefore the same as that proposed by Doncaster *et al.*² — namely, that competition within a fixed set of niches can provide an advantage to sex, with "the success of the clone [being] restrained by the narrowness of its ecological range" and leading to "a stable equilibrium

L. McCarroll*, M. G. Paton*, S. H. P. P.

Karunaratne†, H. T. R. Jayasuriya‡,

K. S. P. Kalpage§, J. Hemingway*

*School of Biosciences, Cardiff University,

Cardiff CF10 3TL, UK

e-mail: Hemingway@cardiff.ac.uk

†Zoology Department, Peradeniya University,

Peradeniya, Sri Lanka

‡Zoology Department, Open University, Nawala,

Nugegoda, Sri Lanka

§Anti-Filaria Campaign, Colombo, Sri Lanka

- Hemingway, J. & Ranson, H. *Annu. Rev. Entomol.* **45**, 369–389 (2000).
- Hemingway, J. & Karunaratne, S. H. P. *Med. Vet. Entomol.* **12**, 1–12 (1998).
- Raymond, M., Callaghan, A., Fort, P. & Pasteur, N. *Nature* **350**, 151–153 (1991).
- Hemingway, J. *et al. Ins. Mol. Biol.* **9**, 93–99 (2000).
- Paton, M. G. *et al. Biochem. J.* **346**, 17–24 (2000).
- Hemingway, J. *Ins. Biochem. Mol. Biol.* **30**, 1009–1015 (2000).
- Jayasekera, N., Kalpage, K. S. P. & De, S. *Trans. R. Soc. Trop. Med. Hyg.* **85**, 250–254 (1991).
- Peiris, H. T. R. & Hemingway, J. *Med. Vet. Entomol.* **10**, 283–287 (1996).
- Nicolas, L. & Plichart, C. *Parasite* **4**, 253–257 (1997).
- Salazar, C. E. *et al. Ins. Mol. Biol.* **3**, 1–13 (1994).
- Vaughan, A., Hawkes, N. & Hemingway, J. *Biochem. J.* **325**, 359–365 (1997).
- Brogdon, W. G. & Barber, A. M. *Pest. Biochem. Physiol.* **37**, 130–139 (1990).
- Karunaratne, S. H. P. *Southeast Asian J. Trop. Med. Pub. Hlth* **30**, 460–469 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

at which both sexual and clonal individuals persist"⁴.

Although the mechanisms favouring sex appear to be identical, the elegantly simple form of Doncaster *et al.*'s Lotka–Volterra model means that it requires implicit assumptions and is therefore hard to compare with previous tangled bank models, which were based on numerical simulations and which made more explicit assumptions^{4,5}. For example, does Doncaster *et al.*'s model require competition between siblings (as with some formulations of the tangled bank), or is it the special case in which sib competition is excluded³?

It is hard to test the implicit assumptions in the new model², but it may render some aspects of the tangled bank mechanism more testable by emphasizing two parameters: the maximum population growth rate, R_0 , and the degree of overlap between sexual and asexual niches, α . However, the importance of these parameters has been discussed previously — for example, it has been pointed out⁶ that, analogous to variation in R_0 , higher fecundity leads to greater sib competition and so increases the advantage of sex. It has also been shown⁴ how the advantage of sex varies with a 'competition coefficient', which is a version of α . Furthermore, such competition coefficients are notoriously difficult to measure⁷.

Second, does Doncaster *et al.*'s model aid our understanding of the patterns of sexuality observed in nature? Several theoretical problems have been pointed out with the tangled bank model, and empirical studies have tested its (identical) assumptions and predictions.

For example, the advantage of sex is reduced if asexual populations are made up of multiple clones that fill multiple portions of the niche space^{4,5}; coexistence of sexuals and asexuals is rarely observed⁴; plant and algae populations consisting of mixtures of genotypes have carrying capacities not much greater than that of the average of their component genotypes, and rarely greater than that of the best genotype⁸; and results from natural populations indicate that the ecological and demographic predictions of the tangled bank model are not met^{6,9}.

Third, does the model of Doncaster *et al.* provide new insights into the cost of sex (males)? The authors argue that the cost of males is ecology dependent, so there is not necessarily a twofold advantage to be recouped in adaptive payoffs. However, one of the model's implicit assumptions is that different genotypes use different niches (genotype-by-environment interactions), so the model is not purely ecological.

The way in which this type of model allows coexistence between sexuals and asexuals has already been discussed^{4,5}. Moreover, this mechanism is more appropriately viewed as an adaptive payoff in its own right, as in previous formulations^{1,3–5}. In this case, it may reduce the cost to be paid by other mechanisms, but such interactions between models have been discussed extensively¹⁰.

Nonetheless, Doncaster *et al.* remind us of the role that the tangled bank mechanism could play in a pluralist explanation of sex, possibly by interacting with deleterious mutations in a fashion analogous to the Red Queen¹⁰.

S. A. West, A. D. Peters

Institute of Cell, Animal, and Population

Biology, University of Edinburgh,

Ashworth Laboratories, West Mains Road,

Edinburgh EH9 3JT, UK

e-mail: andy.peters@ed.ac.uk

- Kondrashov, A. S. *J. Hered.* **84**, 372–387 (1993).
- Doncaster, C. P., Pound, G. E. & Cox, S. J. *Nature* **404**, 281–285 (2000).
- Ghiselin, M. T. *The Economy of Nature and the Evolution of Sex* (Univ. California Press, Berkeley, 1974).
- Bell, G. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Univ. California Press, Berkeley, 1982).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
- Burt, A. & Bell, G. *Nature* **326**, 803–805 (1987).
- Begon, M., Harper, J. L. & Townsend, C. R. *Ecology* 3rd edn (Blackwell Science, Oxford, 1996).
- Bell, G. *Selection: The Mechanism of Evolution* (Chapman & Hall, New York, 1997).
- Lively, C. M. *Nature* **328**, 519–521 (1987).
- West, S. A., Lively, C. M. & Read, A. F. *J. Evol. Biol.* **12**, 1003–1012 (1999).

Regenerating the damaged central nervous system

Philip J. Horner & Fred H. Gage

The Laboratory of Genetics, 10010 North Torrey Pines Road, The Salk Institute, La Jolla, California 92037, USA

It is self-evident that the adult mammalian brain and spinal cord do not regenerate after injury, but recent discoveries have forced a reconsideration of this accepted principle. Advances in our understanding of how the brain develops have provided a rough blueprint for how we may bring about regeneration in the damaged brain. Studies in developmental neurobiology, intracellular signalling and neuroimmunology are bringing the regeneration field closer to success. Notwithstanding these advances, clear and indisputable evidence for adult functional regeneration remains to be shown.

Injury to the adult central nervous system (CNS) is devastating because of the inability of central neurons to regenerate correct axonal and dendritic connections. The consequences of injury are not just a break in communication between healthy neurons, but a cascade of events that can lead to neuronal degeneration and cell death (Box 1). In contrast to fish, amphibia, and the mammalian peripheral nerves and developing central nerves, adult central mammalian neurons do not regrow functional axons after damage. The inability of the adult neurons to regrow after injury cannot be entirely attributed to intrinsic differences between adult CNS and all other neurons. As reported by Ramón y Cajal¹ in 1928, Tello showed in 1911 that adult CNS neurons could regrow if they were provided access to the permissive environment of a conditioned sciatic nerve. Seventy years passed before Aguayo and colleagues² replicated these studies with new methods that definitively confirmed that adult CNS neurons have regenerative capabilities. This finding revealed that the failure of CNS neurons to regenerate was not an intrinsic deficit of the neuron, but rather a characteristic feature of the damaged environment that either did not support or prevented regeneration. In the past 20 years, progress has been made in identifying the elements that are responsible for the differences between the adult CNS and peripheral nervous system environments, and in the past few years the molecular and cellular bases of regenerative compared with non-regenerative responses are beginning to be revealed.

Regeneration strategies developed from these new discoveries will be applicable to many CNS disorders. Spinal cord and, to a lesser degree, brain injury could be the most approachable, owing to the well defined loss of cells and axons and the relative lack of chronic pathological sequelae. Genetic disorders that result in aberrant axonal pathfinding (such as MASA (mental retardation, aphasia, shuffling gait and adducted thumbs) or Kallmann syndrome) or neuronal cell loss (such as macular degeneration or retinitis pigmentosa) may also be amenable to regeneration. Degenerative diseases where a defined cell phenotype is lost, such as Parkinson's disease, Alzheimer's disease or amyotrophic lateral sclerosis, are also good targets, but may be more challenging because of the potential for continued cell loss or axonal degeneration. Finally, regeneration strategies may also be applied to less well defined disorders where diffuse cell and axonal loss can occur, such as vascular disease, tumour and infection of the CNS.

Strategies for regenerating the adult CNS

Regeneration in the adult CNS requires a multistep process. First, the injured neuron must survive, and then the damaged axon must extend its cut processes to its original neuronal targets. Once contact is made, the axon needs to be remyelinated and functional synapses

need to form on the surface of the targeted neurons. Several strategies have been undertaken that target different aspects of this process. We will summarize and evaluate critically the latest data emerging from the following regeneration strategies: cellular

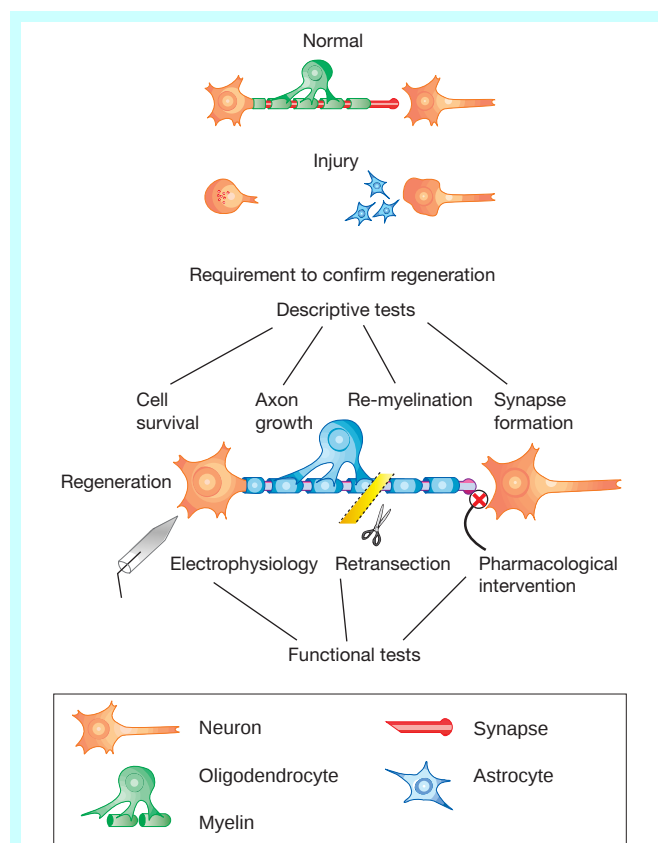
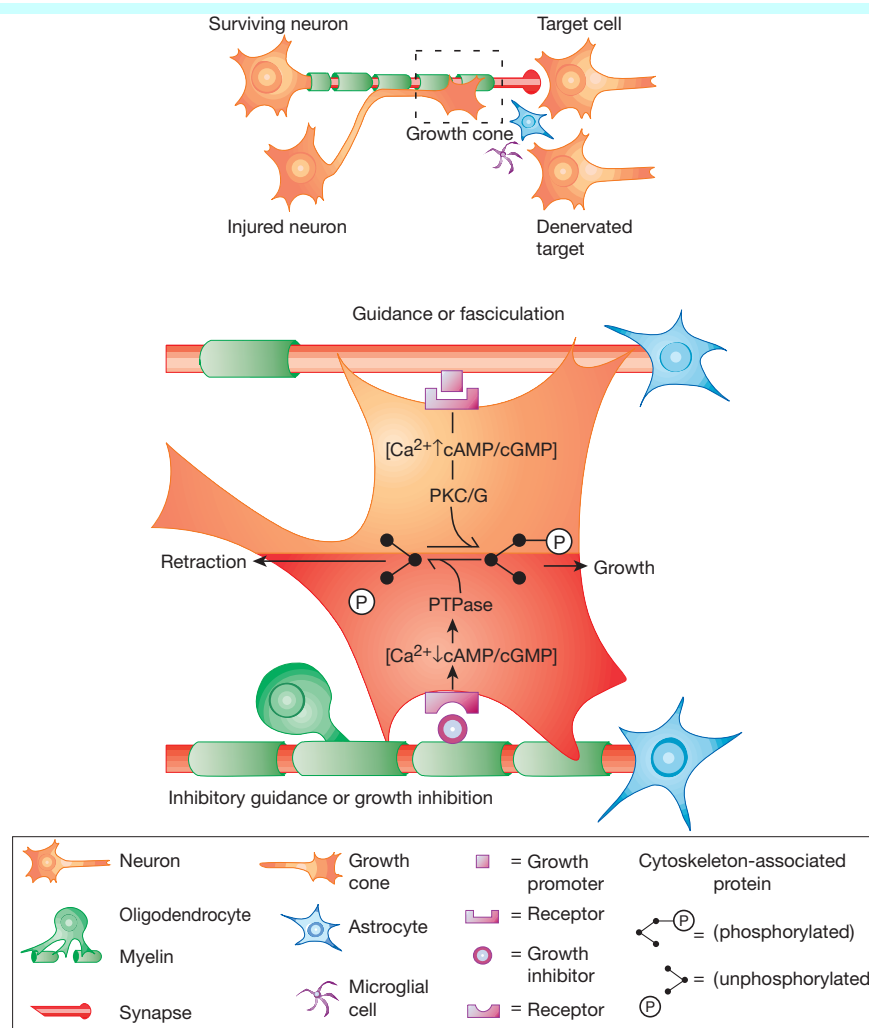


Figure 1 Steps to functional regeneration. Several criteria need to be met before functional regeneration can be validated. A combination of descriptive and functional analyses needs to be used to critically evaluate an experiment. Descriptive tests can be used to determine the survival and integrity of the injured system, whether axonal regeneration is present, and if appropriate synaptic connections and remyelination have occurred. The goal is to correlate the descriptive data with direct physiological and behavioural evidence for regeneration. Electrophysiological and pharmacological intervention can be used to assess the function and specificity of the regenerated pathway. Ultimately, elimination of the regenerated pathway (for example, retranssection) is important to determine its role in any reported functional recovery.

Figure 2 Intracellular mediators of axon growth. Regeneration in the CNS requires that axons migrate long distances through the adult CNS to reconnect with an appropriate target. The axon growth cone contains the machinery for growth and must navigate the intact and injured CNS to reconnect with a target. Along its path, a growth cone can be diverted or converted to a non-migratory retraction bulb when it encounters inhibitory factors. Inhibition can be induced by molecules expressed on the surface of reactive cells and exported to the extracellular matrix, by inhibitory molecules found on myelin and by lack of diffusion or tropic molecules. Axonal growth and fasciculation can be induced by molecules expressed on other regenerating axons or by surviving axons, and by supportive molecules expressed by reactive cells. Many of these effectors appear to feed into the growth cone through a common mechanism involving the regulation of intracellular messengers. Changes in the levels of intracellular messengers affect the phosphorylation state of cytoskeleton-associated proteins, which leads to axon growth or collapse. Modification of these intracellular mediators may be a powerful tool for converting inhibition to attraction within the injured CNS. PKC, protein kinase C.



replacement, neurotrophic factor delivery, axon guidance and removal of growth inhibition, manipulation of intracellular signaling, bridging and artificial substrates, and modulation of the immune response.

As described below, an increasing number of studies have shown that an adult cut axon can be induced to regrow by either increasing the permissive cues or decreasing the non-permissive cues of the existing environment. Furthermore, a growing list of reports show that one strategy or another can induce some level of functional recovery following damage. However, it is not sufficient to demonstrate axon elongation and behavioural improvement after injury to conclude that authentic functional regeneration is responsible for the outcome. There are many mechanisms that may account for observed functional recovery that do not require regeneration. These non-regenerative mechanisms are common in most experimental models of traumatic injury and need to be excluded before invoking functional regeneration as the cause of repair and recovery. The reason for sorting out the authentic mechanisms of functional recovery is that, without understanding the underlying basis of regeneration, little progress can be made beyond the phenomenological observation of recovery from injury.

Cellular replacement

In most cases of CNS trauma and disease, both neurons and glia are lost. Cell replacement is a vital step in CNS lesions, in which spared systems cannot supplant the function of lost cells. An excellent example is the loss of motor neurons in the case of injury to the

cervical spine. No matter how well the lesion is bridged or re-innervated, the target motor neurons that project to the muscles are irrevocably lost. Recognizing the need for cell replacement, several laboratories have taken advantage of fetal tissue grafts to replace cells after a variety of CNS insults³. These studies have even led to clinical trials for humans. However, the inherent mechanical, physical and ethical hurdles of using fetal tissue limit its large-scale use. Ideally, a cell replacement source should be more homogeneous, readily obtainable and syngeneic.

Stem cells have great promise as a source for introducing new neurons or glial cells to the damaged CNS. Neural stem cells have now been isolated from diverse regions of the developing and adult rodent brain and cultured using either epidermal or fibroblast growth factor⁴. Stem cells are advantageous for research and potential therapy because they are multipotent, and can be propagated *in vitro*, genetically tagged with markers or therapeutic genes, and grafted into the developing and mature intact CNS⁵⁻⁷. Once transplanted into the brain, neural stem cells can adapt to the region of engraftment by differentiating into the appropriate neuronal and glial subpopulations⁸. In an extreme example, a single study has reported that cells derived from the adult brain can become cells of the blood lineage when transplanted into irradiated mice⁹. Replication of the study has not been reported and is necessary if we are to consider this potential further. In addition, several studies have reported that stem-cell grafts can lead to neurogenesis in animal models of degeneration. For example, cortical neurons undergoing damage-induced apoptosis can be replaced morphologically by

transplanted embryonic neural stem cells that morphologically resemble pyramidal cells¹⁰. Replication and subsequent tests are necessary to determine whether grafted cells are functional.

Evaluating functional replacement. Despite encouraging data that stem cells can morphologically replace cells of the CNS, few studies have shown a functional impact for this approach. One exception is a study showing the replacement of oligodendrocytes by transplanted embryonic stem (ES) cells in a spinal-cord-injured rat¹¹. The result was improved locomotion; however, it is not clear whether the ES cells that survived contributed to the structural reorganization following injury, or whether the cells secreted some factor that influenced the remaining damaged tissue. Although it is still unclear how these transplants may work, they illustrate an important point, specifically that neural replacement cells may not necessarily need to be derived from the brain. For example, ES cells and bone marrow stromal cells have been transplanted to the adult brain, where they are reported to differentiate into neural cells^{12–14}. This is promising, but the criteria to determine that a surviving neural cell is differentiated and functional need to become more rigorous as this field advances.

Stimulation of self repair. The presence of neural stem cells resident within the CNS and the ability to regulate their numbers and fate may provide an alternative to transplantation. The existence of neural stem cells has recently been shown in non-human primates and humans^{15,16}. In addition, the adult brain⁴ and spinal cord¹⁷ have been shown to contain stem cells that continually divide throughout adult life. Recent studies have shown that cell genesis and the synaptic plasticity of these cells can be influenced by stress, an enriched environment and physical exercise^{18–20}. With respect to injury, progenitor cells are capable of proliferation and differentiation into mature myelinating oligodendrocytes^{21–23} in models of acute demyelination. In addition, cortical neurons are replaced by dividing progenitor cells in a model of selective pyramidal cell apoptosis²⁴. Strikingly, these newborn neurons extend new axons several millimetres through the intact CNS. These findings show that under the correct lesion conditions, CNS stem cells are capable of participating in cell replacement. Despite their inherent plasticity, however, it is clear that endogenous stem cells do not produce complete recovery in cases of severe trauma.

Whether a strategy of transplantation or endogenous replacement is used, the molecules and genes that govern cell proliferation, migration and differentiation need to be determined. In addition, new cells that are grafted or generated *in situ* will need to be engineered to be resistant to, or pharmacologically protected from, the toxic environment of the injured CNS. Finally, even when dead cells can be replaced with new ones, the new cells will need to be functionally integrated into the remaining circuitry, perhaps by surviving cells or by externally driven programs.

Neurotrophic factor delivery

Many studies have shown both the cell survival and axon growth-promoting effects of neurotrophic molecules after injury to the adult CNS *in vivo*²⁵. Neurotrophins are not typically used as a singular effector, however, but rather are combined with growth-promoting cells or matrices²⁶. The assumption is that neurotrophins will induce axonal growth only when the appropriate growth-permissive substrate is present. Schwann cells provide a permissive substrate for propriospinal, sensory and supraspinal axons, although axons do not then re-enter the CNS²⁷. For example, Menei *et al.*²⁸ used engineered Schwann cells that express brain-derived neurotrophic factor (BDNF) to regrow supraspinal axons across a transected adult spinal cord. Schwann cells were seeded into a polymer channel that was placed in between the spinal cord stumps, and a Schwann cell trail was created by injecting engineered cells into the proximal and distal stump. In this experiment, BDNF did provide increased selection for trkB-expressing axons, which demonstrates the utility of neurotrophin selection; however, axons

once again did not leave the permissive substrate of the guidance channel or Schwann cell trails.

Are neurotrophins sufficient? These studies illustrate one of the principal obstacles to the use of neurotrophins: the need to drive regenerating axons out of a permissive substrate or bridge and into the injured CNS. Some studies have shown that experimentally induced increases in neurotrophin levels within the tissue may be enough to facilitate this process. Grill *et al.*²⁹ used NT-3-producing fibroblasts to stimulate the regeneration of the cortical spinal tract in the hemisectioned adult rat spinal cord. In a similar study, Liu *et al.*³⁰ used BDNF-producing fibroblasts to increase the growth of rubrospinal axons in the hemisectioned spinal cord. In both of these studies, tract tracing showed axons that regenerated around the lesion site and several millimetres into the target region, with a correlated improvement in behaviour. These findings suggest that neurotrophins alone can stimulate axonal migration through the injured CNS, and that regeneration beyond the injury site can be associated with functional improvements. However, these studies do not prove a direct relationship between axonal regeneration and behavioural outcome. Alternative explanations include axonal sprouting of non-injured axons, activation of redundant pathways and alterations in the receptor number/phenotype or excitability of surviving neurons or glia. The existence of these mechanisms is well demonstrated in three studies in which behavioural changes were documented to occur as a result of neurotrophin administration, but no evidence for long-tract regeneration was found^{31–33}.

Evaluation of functional regeneration. The findings of functional changes without substantial regeneration argue for a more detailed analysis of reported axonal regeneration (Fig. 1). One recent example comes from Ramer *et al.*³⁴, who showed that phenotypically appropriate regeneration of sensory axons into the CNS could be accomplished by pumping neurotrophins intrathecally. This study is particularly thorough because, in addition to anatomical tracing, the investigators used both specific behavioural tests corresponding to the regenerated fibre phenotypes and electrophysiology for confirmation of connectivity. Most importantly, they subsequently re-injured the regenerated fibres to demonstrate their requirement for functional recovery. Extension of these types of analyses to injured central axons is needed.

Despite the extensive use of neurotrophins, there are many remaining challenges. Neurotrophic factors comprise a large and complex family of molecules, and many subclasses of neurons and glia express unique complements of neurotrophic factor receptors. Many questions remain regarding practical application of these molecules; for example, diffusion of exogenously delivered neurotrophins is variable. In addition, neurotrophins have many properties aside from their roles in neuronal survival and axonal growth. For example, neurotrophins are anterogradely transported and released from presynaptic to postsynaptic targets. Neurotrophins also modulate membrane excitability, induce neuronal hypertrophy, affect cell differentiation, and result in broad systemic effects. These functions for neurotrophins will need to be considered when interpreting future studies.

Axon guidance and removal of growth inhibition

Several growth-promoting molecules have a developmental role in axon guidance, fasciculation, synapse formation and, in the intact or injured adult, regeneration and activity-dependent plasticity (Table 1). Although the existence of axon guidance molecules can be inferred from transplantation experiments³⁵, few direct data exist to show that specific growth-promoting molecules influence regeneration in the adult CNS. Notable exceptions include L1 and polysialic acid neural cell adhesion molecule (PSA-NCAM), which have been shown to be associated with regenerating axons in the hippocampus or spinal cord-lesioned rat^{36,37}.

Identification of axon guidance factors on a molecular level has been a relatively recent event and, as a result, the use of these factors

has been limited in studies of repair. With the knowledge that adult central neurons can grow when given an alternative to the CNS environment, considerable effort has been placed on identifying and reversing the effects of growth inhibitory molecules. Proof that the CNS contains actual growth inhibitors was first provided by Caroni, Schwab and Savio^{38,39} in 1988. They discovered that the inhibitory nature of cultured oligodendrocytes could be blocked by the monoclonal antibody, IN-1. Recently, the gene for this inhibitor has been sequenced and named *nogo*⁴⁰. Since Caroni and Schwab's original discovery, several other factors have been identified that can inhibit axon growth, including a variety of glycoproteins and proteoglycans^{41,42}. In addition, the concept of axonal guidance by inhibitory molecules was introduced from studies in the developing nervous system^{43–45}.

Inhibition by the gliotic scar. Many of the inhibitory factors identified for the immature and adult brain may be re-expressed after CNS injury in the adult. The chemorepellant semaphorin III/collapsin I is re-expressed in scar tissue after injury to several areas of the adult CNS⁴⁶. The axon growth inhibitors chondroitin sulphate and keratin are re-expressed after injury to the mature brain^{47,48} and spinal cord⁴⁹. The proteoglycan NG2, thought to be expressed on immature glial cells, is increased after cortical injury and can inhibit axon growth when produced by astrocytes⁵⁰. Interestingly, an important characterization of the astrocytic response to injury has revealed heterogeneity in the expression of inhibitors by glial cells⁵¹. This finding underscores the complexity of the environment through which an injured axon must regrow and may reveal opportunities for intervention. Another important finding is the physical evidence that glial scarring can inhibit diffusion between cells at the injury site. Calculating diffusion coefficients with tetramethyl ammonium-sensitive electrodes, Roitbak and Sykova⁵² have shown that diffusion is decreased in regions of astrocytic hypertrophy and increased chondroitin sulphate. This finding indicates that scar may not only contain molecular inhibitors but may also act as a simple barrier to the diffusion of growth-promoting molecules.

The contribution of scar formation to axonal inhibition has been tested by using chondroitinase to increase the permissiveness of injured spinal cord sections to embryonic dorsal root ganglion axons⁵³. Similarly, deletion of scar-forming astrocytes by gene targeting results in increased axonal sprouting accompanied by a dysfunctional blood-brain barrier and increased immune activation *in vivo*⁵⁴. Work by Stichel *et al.*⁵⁵ has suggested that the impermeable nature of scar is due to its basal membrane. Reduction of collagen type IV in the lesioned fornix resulted in axonal migration in the scar region despite continued expression of proteoglycans and glial activation. However, this approach did not improve growth of the injured cortico-spinal tract⁵⁷.

Inhibition by adult myelin. Considerable attention has been directed towards the disruption of myelin inhibitors. The addition of monoclonal antibodies to the neurite inhibitor IN-1 alone resulted in axonal regeneration of the cortical spinal tract and functional return of forelimb use in the adult rat⁵⁶. Disruption of myelin by immune activation⁵⁷ or induction of an autoimmune reaction to myelin⁵⁸ also enhanced regeneration in the adult spinal cord. These data illustrate that blocking or eliminating endogenous white matter inhibitors may permit the natural regeneration potential of injured nerves. However, the role of myelin inhibitors has been challenged by a study in which micro-injected adult peripheral sensory neurons were capable of extending axons along myelin tracts in the intact adult rat spinal cord⁵⁹. These data indicate that, although myelin inhibitors exist, the geometry of white matter may be such that myelin is not always inhibitory and can even, in this case, be permissive. The concept of myelin geometry was further addressed in an *in vitro* model of regeneration where peripheral axons were inhibited only when the alignment of white matter was perpendicular to the direction of axonal growth⁶⁰. It is vital to keep

in mind that these two experiments examined peripheral sensory neurons, and their behaviour may differ from that of neuronal cells that reside within the CNS.

Collectively these data indicate that growth inhibition may be caused by molecules that are present in the uninjured state, but also by production of the glial scar, which represents a molecular and physical barrier to regeneration. However, it is important to consider the novel role of growth inhibitory molecules in guiding axons along a white matter tract or around a cyst that no longer contains an appropriate target or scaffold. In the future, we will undoubtedly see increased manipulation of axon guidance molecules, both inhibitory and permissive, as a method for circumventing scar and stimulating axon growth.

Manipulation of intracellular signalling

Potential mediators of cell and axon fate. There are many diffusible and substrate-bound factors vital to cell survival and axonal growth. The intracellular mediators of these effects are beginning to be understood. One potential method for preventing cell death and stimulating regeneration is to alter the intracellular signals that a cell uses to transduce responses to injurious or growth-inhibiting substrates. For example, in apoptotic cell death, several transcription factors (such as *fas*, *p53*, *c-Jun*, *bax* and *bcl-2*) are involved and may represent common pathways to intervene and prevent cell loss⁶¹. Apoptotic cell death has been inhibited by blocking the ionic changes that occur near the cell, increasing protease inhibitors that block the downstream effectors of cell death signals⁶² or upregulating anti-apoptotic genes⁶³. With regards to axon growth, several intracellular factors, including growth-associated proteins, Ca^{2+} concentrations, transcription factors, and second messengers such as the cyclic nucleotides and inositol triphosphate, appear to have key roles (Fig. 2). Growth-associated proteins such as tubulin, actin and GAP-43 are thought to be harbingers of critical growth periods after injury and during development^{64–67}. Manipulation of their expression may be important for stimulating a cell to re-enter a growth mode. The anti-apoptotic transcription factor *bcl-2* also plays a role in axonal plasticity⁶⁸ and regeneration following injury⁶⁹. Although high intracellular Ca^{2+} has long been thought to be a mediator of cell death, recent data suggest that frequent Ca^{2+} transients within a growth cone are correlated with retarded growth⁷⁰. Several reports suggest that the intracellular concentrations of cyclic AMP and cyclic GMP can modulate the response of a growth cone to inhibitory and stimulatory factors found in its environment^{71,72}. In general, it appears that raising intracellular cyclic nucleotide concentrations is sufficient to convert an inhibitory response to an attractive one. For example, axonal repulsion of *Xenopus* spinal axons induced by semaphorin III/collapsin-1 is reversed in the presence of an analogue of cGMP⁷³. The next step is for these intriguing *in vitro* findings to be translated to *in vivo* contexts.

Application of signalling molecules *in vivo*. Modification of intracellular responses has the benefit of generating a broad generalized response from the cell. This approach may be appealing for cell preservation and regeneration, as it may eliminate the need for multiple inhibitors of the many factors shown to induce apoptosis and growth cone collapse. However, there are several concerns about *in vivo* application. Systemic approaches do not take into account the variety of cell types that exist in the CNS and their individual responses to injury. In addition, Ca^{2+} channel blockers and cyclic nucleotide analogues have adverse systemic side effects and can be lethal. Finally, timing is probably a vital issue. A recent study suggests that, once a sprouting axon has encountered inhibitory substrates, it loses its ability to respond to neurotrophins and their intracellular mediators⁷⁴.

Bridging and artificial substrates

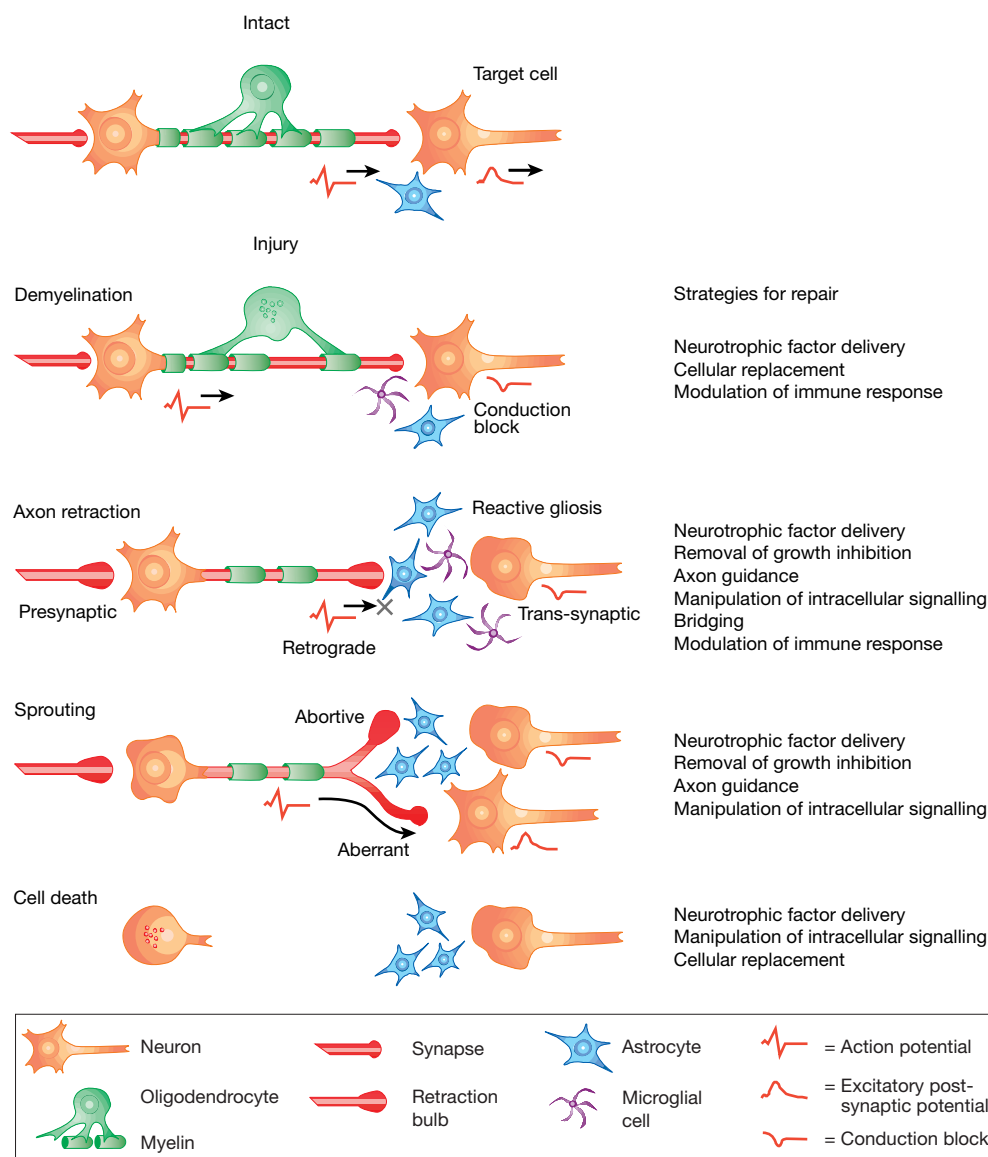
Cellular bridging. The concept of bridging is most apparent in

lesions of the CNS where a portion of the CNS is lost, a cyst has formed or regions of extensive scarring develop. A bridge is used to guide axons across these barriers and re-introduce axons into the remaining intact parenchyma. This familiar concept has led to some of the most promising evidence for regeneration using peripheral nerves or fetal tissue in the visual system and spinal cord. Several

new approaches are being developed that may complement the task of collecting peripheral nerves or embryonic tissue. However, the important advances will be guided by our molecular understanding of how the Schwann cell or the immature nervous system provides a permissive substrate for axonal growth. Genetic engineering of skin fibroblasts has shown that overexpression of neurotrophins⁷⁵ can

Box 1

Consequences of neuronal injury and strategies for repair



Following a specific traumatic or chemotoxic event, or as a result of ongoing degenerative processes, long-term structural and functional deficits occur in the adult CNS. In severe cases, these insults are not repaired or compensated for by surviving systems. On a cellular level, these deficits include demyelination, degeneration, abortive or aberrant sprouting, and cell death.

Demyelination. A demyelinated axon may maintain both its afferent and efferent connections but, due to a loss of myelination, poor or failed conduction results.

Axonal retraction. Injury to an axon itself or to the original cellular target of the axon can result in degeneration. Presynaptic, retrograde and trans-synaptic degeneration can occur. Synaptic conduction across a pathway is lost and a reactive cellular response, including

astrocytes and microglia, forms.

Sprouting. Axonal sprouting has been described for surviving neurons. It is typically abortive when a sprouting axon encounters an inhibitory matrix or scar, loss of neurotrophin support, or the presence of continuing inflammation or toxicity. Aberrant sprouting can occur when an axon reconnects to an inappropriate target. Synaptic conduction is restored but this pathway does not result in functional restoration.

Cell death. When a neuron is completely deprived of its source of growth factors and exposed to high levels of toxic molecules or inflammatory attack, it can undergo cell death. These patterns represent the anatomical correlates of brain dysfunction but also the specific processes that must be targeted for repair.

make these cells more permissive for axonal regeneration. In addition, cultured Schwann cells have been used to replace whole peripheral nerves^{37,76}. Type I collagen has been used as a substrate for grafting cells⁷⁷, or alone as a bridge mixed with neurotrophins³². Several recent reports have shown that cultured ensheathing glial cells taken from the olfactory epithelium may have the unique ability to guide injured axons back into the intact CNS^{78–80}. In each case, ensheathing glia grafted into the injured spinal cord show extensive migration and facilitate the regeneration of axons. These qualities are somehow tied to their intrinsic role of guiding and ensheathing growing axons across the border between the CNS and the peripheral nervous system in the adult olfactory bulb. As the mechanism by which ensheathing glial cells promote axonal regrowth remains to be reported, a call for a more physiological and molecular understanding of this process is warranted.

Artificial substrates. In the future, artificial substrates may be useful for repair of lesions where bridging is necessary. The ideal artificial substrate will have a molecular make-up that is easily manipulated and immune tolerant, and will contain a porous scaffold for nerve regeneration and cell repopulation that will be easily absorbed by the CNS. In the transected spinal cord, regeneration has been guided by artificial tubes that are seeded with growth-promoting Schwann cells^{27,81}. These are composed of acrylonitrile and vinylchloride and are semi-rigid, porous tubes that can be implanted and survive indefinitely. Remaining concerns include mechanical hindrance and lack of absorption. Rapidly absorbable substrates are being tested, such as poly- α -hydroxyacids in the transected spinal cord⁸². Freeze-dried alginate gels and gel foam soaked in neurotrophins are reabsorbable materials that have been

used to promote cell survival and regeneration⁸³. One of the primary limitations of artificial substrates is that invasive surgery is often necessary to install them. This downside is particularly relevant to the CNS. Although there are many hurdles to overcome, reabsorbable substrates that are integrated with cells, guidance molecules or electrical fields represent an exciting new approach for repair.

Modulation of the immune response

Does the immune system exacerbate injury? The immune system's role in injury and the process of regeneration involves peripheral and central, and cellular and humoral immune responses⁸⁴. For a considerable time there has been speculation about dualism of the immune response to injury in the brain. On the positive side, several investigators have suggested that the immune system may be protective and even assist regeneration. On the negative side, the immune system may cause bystander damage and progressive necrosis during the process of eliminating dead or dying tissue. Recently, investigators have begun to dissect out detrimental aspects of the immune system that may be responsible for unnecessary tissue loss and restrictions on axonal regeneration. For example, early inhibition of tumour necrosis factors or transforming growth factor- β 2 significantly decreases scarring and tissue loss, and can lead to improvement in functional outcome after CNS injury^{85,86}. Application of interferon- γ modulates several extracellular matrix molecules found in the CNS scar and inhibits astrocyte proliferation⁸⁷. These observations demonstrate that certain aspects of the immune response may substantially restrict axonal plasticity following trauma.

Regeneration and the central immune response. Blinzinger and

Table 1 Molecules involved in axon growth and guidance

Family	Ligands	Receptors*	Cellular expression†	Functional state‡	General description
Extracellular matrix molecules	Laminin, tenascin, collagen, fibronectin, immunoglobins, integrins, anosmin-1 (KAL), chondroitin/heparan sulphate proteoglycans, thrombospondin-1	HomA, proteoglycans, integrins	F, A, NE, N	ECM	Components of the ECM that bind to a number of membrane-bound receptors. Can be permissive or non-permissive. Many important in neuronal development. All are expressed in the adult CNS.
Semaphorins/collapsins	Sema 3A-F, Sema 4A-G, Sema 5A and B, Sema 6A-C, Sema 7A	Plexin A, neuropilin	N, A, Ax	RD, IM	Chemorepellents that stimulate growth cone collapse in sensory axons <i>in vivo</i> . Some forms may act as an attractant. Can exist as integral membrane proteins or secreted.
Immunoglobins	N-CAM, PSA-N-CAM, L1, TAG-1/axonin-1, DM-Grasp, cadherins	HetA, HomA	Ax	IM, ECM	Integral and membrane-bound glycoproteins involved in axonal attachment, fasciculation and guidance. Some exhibit homophilic and heterophilic interactions between self and other family members.
Myelin-associated inhibitors	Myelin-associated glycoprotein (MAG), nogo	Unknown	O	IM	Integral membrane proteins expressed in adult myelin and are inhibitory <i>in vitro</i> and <i>in vivo</i> . May act as receptors for certain ECMs.
Tyrosine kinase receptors	Ephrin A1–A5, Ephrin B1–B3	Epha A1–A8, Eph B1–B6	Ax, N	IM	Ligand and receptors are membrane-bound or transmembrane proteins involved in cell and axonal migration during development.
Netrins	Netrin 1,2	Deleted in colorectal cancer (DCC), neurogenin, unc-5	N, Ax, NE	RD, ECM	Important in neuronal and axonal migration in the developing CNS. Interactions with the DCC receptors can be either chemorepellent or attractive.
Neurotrophic factors	NGF, NT-3, BDNF, CNTF, GDNF	Tyrosine-kinase receptor A,B,C, p75, LfR β , GFR α 1	NE, N, Ax, A, O, M, F	RD	Neurotropic and neurotrophic capabilities in the injured adult CNS <i>in vivo</i> and <i>in vitro</i> .
Growth factors	FGF, IGF, PDGF, VEGFa	FGFR α 1–4, flk-1, IGF α 1, PDGFR, EGFR	NE, N, Ax, A, O, M, F	RD, ECM	Peptide growth factors involved in cell proliferation, differentiation and cell death. <i>In vitro</i> and <i>in vivo</i> data support a role for neurotrophic activity.
Inflammatory cytokines	TGF- β , LIF, TNF, EGF	TGF- β type I, TGF- β type II, LIFR β , gp130, TNFR α 1–2	M, A	RD	Peptide growth factors that modulate the immune system, but also have been shown to have a role in neuronal cell differentiation and neurotrophic activity <i>in vitro</i> .
Neurotransmitters	Acetylcholine (ACh)	AChR	N, Ax	RD	Possibly involved in synapse formation during development. <i>In vivo</i> data indicate that ACh induces turning in sensory growth cones.

* HetA, heterophilic adhesion; HomA, homophilic adhesion.

† A, astrocyte; Ax, axon; F, fibroblast; M, microglia; NE, neuroepithelial; N, neuron; O, oligodendrocyte.

‡ ECM, extracellular matrix; IM, integral membrane; RD, released/diffusible.

Kreutzburg⁸⁸ originally suggested a proregenerative role for the central immune system when they described the microglial response after injury to the facial nerve. Microglial cells are found to contact the membrane of neurons that will not undergo apoptosis and will eventually regenerate their axons. Because this lesion does not involve the peripheral immune response, these anatomical data suggested that microglia may somehow orchestrate the recovery of these cells by assuming a protective role. Supportive cytokines are released after injury to the facial nerve lesion or spinal cord, which suggests that candidate cytokines are involved in immune-mediated repair⁸⁹. Finally, depletion of peripheral macrophages results in tissue sparing and increased neuronal sprouting following spinal cord injury⁹⁰. These data further indicate a disparity between central and peripheral immune responses to injury. In the future, this field will need to continue to identify the pathways that lead to the beneficial and harmful immune responses.

Conclusions

The inherent complexity of the adult CNS as well as its amazing plasticity and adaptiveness make interpreting *in vivo* regeneration experiments challenging. As reports of functional regeneration become more common, it is important to critically evaluate the mechanisms of the observed functional changes. Case by case, it will be important to consider the inherent plasticity and variability of sparing in each injury model. In addition, several questions should be asked when evaluating a regeneration experiment. To what degree does sprouting or compensation by spared systems contribute to recovery? Do anatomical changes occur in conjunction with behavioural and physiological recovery? When regenerated pathways are experimentally interrupted, is functional recovery once again lost? Technological advances will lead to more and better tools, allowing us to address these questions and identify the most promising therapeutic approaches. Identification of the molecules that induce or inhibit regeneration will propel us closer to achieving functional regeneration. It is conceivable that most of these molecules will be discovered in the next 5 years and the first rational, functional therapy for regeneration, probably for spinal cord injury, may be in the clinic just 10 years from now.

Despite the progress in the last century of research on regeneration, however, Cajal's flowery decree, as translated by Raoul May, still resonates: "once the development was ended, the founts of growth and regeneration of the axons and dendrites dried up irrevocably. In the adult centres the nerve paths are something fixed, ended and immutable. Everything may die, nothing may be regenerated. It is for the science of the future to change, if possible, this harsh decree".

The decree is lifted; the solution remains elusive. □

1. Ramon y Cajal, S. *Degeneration and Regeneration of the Nervous System* (Hafner, New York, 1928).
2. Richardson, P. M., McGuinness, U. M. & Aguayo, A. J. Axons from CNS neurones regenerate into PNS grafts. *Nature* **284**, 264–265 (1980).
3. Kordower, J. H. & Tuszynski, M. H. *CNS Regeneration: Basic Science and Clinical Advances* (eds Kordower, J. H. & Tuszynski, M. H.) 159–182 (Academic, San Diego, 1999).
4. Gage, F. H. Mammalian neural stem cells. *Science* **287**, 1433–1438 (2000).
5. Flax, J. D. et al. Engraftable human neural stem cells respond to developmental cues, replace neurons, and express foreign genes. *Nature Biotechnol.* **16**, 1033–1038 (1998).
6. Gage, F. H. et al. Survival and differentiation of adult neuronal progenitor cells transplanted to the adult brain. *Proc. Natl Acad. Sci. USA* **92**, 11879–11883 (1995).
7. Liu, Y. et al. Intraspinal delivery of neurotrophin-3 using neural stem cells genetically modified by recombinant retrovirus. *Exp. Neurol.* **158**, 9–26 (1999).
8. Gaiano, N. & Fishell, G. Transplantation as a tool to study progenitors within the vertebrate nervous system. *J. Neurobiol.* **36**, 152–161 (1998).
9. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells *in vivo*. *Science* **283**, 534–537 (1999).
10. Snyder, E. Y., Yoon, C., Flax, J. D. & Macklis, J. D. Multipotent neural precursors can differentiate toward replacement of neurons undergoing targeted apoptotic degeneration in adult mouse neocortex. *Proc. Natl Acad. Sci. USA* **94**, 11663–11668 (1997).
11. McDonald, J. W. et al. Transplanted embryonic stem cells survive, differentiate and promote recovery in injured rat spinal cord. *Nature Med.* **5**, 1410–1412 (1999).
12. Brustle, O. et al. Embryonic stem cell-derived glial precursors: a source of myelinating transplants. *Science* **285**, 754–756 (1999).

13. Ono, K. et al. Migration of exogenous immature hematopoietic cells into adult mouse brain parenchyma under GFP-expressing bone marrow chimera. *Biochem. Biophys. Res. Commun.* **262**, 610–614 (1999).
14. Bartlett, P. F. Pluripotential hemopoietic stem cells in adult mouse brain. *Proc. Natl Acad. Sci. USA* **79**, 2722–2725 (1982).
15. Kornack, D. R. & Rakic, P. Continuation of neurogenesis in the hippocampus of the adult macaque monkey. *Proc. Natl Acad. Sci. USA* **96**, 5768–5773 (1999).
16. Eriksson, P. S. et al. Neurogenesis in the adult human hippocampus. *Nature Med.* **4**, 1313–1317 (1998).
17. Horner, P. J. et al. Proliferation and differentiation of progenitor cells throughout the intact adult rat spinal cord. *J. Neurosci.* **20**, 2218–2228 (2000).
18. Gould, E. & Tanapat, P. Stress and hippocampal neurogenesis. *Biol. Psychiatry* **46**, 1472–1479 (1999).
19. Kempermann, G., Kuhn, H. G. & Gage, F. H. More hippocampal neurons in adult mice living in an enriched environment. *Nature* **386**, 493–495 (1997).
20. van Praag, H., Kempermann, G. & Gage, F. H. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nature Neurosci.* **2**, 266–270 (1999).
21. McTigue, D. M., Horner, P. J., Stokes, B. T. & Gage, F. H. Neurotrophin-3 and brain derived neurotrophic factor induce oligodendrocyte proliferation and myelination of regenerating axons in the contused rat spinal cord. *J. Neurosci.* **18**, 5354–5365 (1998).
22. Franklin, R. J., Gilson, J. M. & Blakemore, W. F. Local recruitment of remyelinating cells in the repair of demyelination in the central nervous system. *J. Neurosci. Res.* **50**, 337–344 (1997).
23. Gensert, J. M. & Goldman, J. E. Endogenous progenitors remyelinate demyelinated axons in the adult CNS. *Neuron* **19**, 197–203 (1997).
24. Magavi, S. S., Leavitt, B. R. & Macklis, J. D. Induction of neurogenesis in the neocortex of adult mice. *Nature* **405**, 951–955 (2000).
25. Stichel, C. C. & Muller, H. W. Experimental strategies to promote axonal regeneration after traumatic central nervous system injury. *Prog. Neurobiol.* **56**, 119–148 (1998).
26. Cheng, H., Cao, Y. & Olson, L. Spinal cord repair in adult paraplegic rats: partial restoration of hind limb function. *Science* **273**, 510–513 (1996).
27. Xu, X. M., Guenard, V., Kleitman, N., Aebischer, P. & Bunge, M. B. A combination of BDNF and NT-3 promotes supraspinal axonal regeneration into Schwann cell grafts in adult rat thoracic spinal cord. *Exp. Neurol.* **134**, 261–272 (1995).
28. Menei, P., Montero-Menei, C., Whittemore, S. R., Bunge, R. P. & Bunge, M. B. Schwann cells genetically modified to secrete human BDNF promote enhanced axonal regrowth across transected adult rat spinal cord. *Eur. J. Neurosci.* **10**, 607–621 (1998).
29. Grill, R., Murai, K., Blesch, A., Gage, F. H. & Tuszynski, M. H. Cellular delivery of neurotrophin-3 promotes corticospinal axonal growth and partial functional recovery after spinal cord injury. *J. Neurosci.* **17**, 5560–5572 (1997).
30. Liu, Y. et al. Transplants of fibroblasts genetically modified to express BDNF promote regeneration of adult rat rubrospinal axons and recovery of forelimb function. *J. Neurosci.* **19**, 4370–4387 (1999).
31. Jakeman, L. B., Wei, P., Guan, Z. & Stokes, B. T. Brain-derived neurotrophic factor stimulates hindlimb stepping and sprouting of cholinergic fibers after spinal cord injury. *Exp. Neurol.* **154**, 170–184 (1998).
32. Houweling, D. A., Lankhorst, A. J., Gispens, W. H., Bar, P. R. & Joosten, E. A. Collagen containing neurotrophin-3 (NT-3) attracts regrowing injured corticospinal axons in the adult rat spinal cord and promotes partial functional recovery. *Exp. Neurol.* **153**, 49–59 (1998).
33. Ribotta, M. G. et al. Activation of locomotion in adult chronic spinal rats is achieved by transplantation of embryonic raphe cells reinnervating a precise lumbar level. *J. Neurosci.* **20**, 5144–5152 (2000).
34. Ramer, M. S., Priestley, J. V. & McMahon, S. B. Functional regeneration of sensory axons into the adult spinal cord. *Nature* **403**, 312–316 (2000).
35. Bahr, M. Target-specific guidance cues for regenerating axons are reexpressed in the lesioned adult mammalian central nervous system. *Adv. Neurol.* **73**, 83–90 (1997).
36. Aubert, I., Ridet, J. L., Schachner, M., Rougon, G. & Gage, F. H. Expression of L1 and PSA during sprouting and regeneration in the adult hippocampal formation. *J. Comp. Neurol.* **399**, 1–19 (1998).
37. Weidner, N., Blesch, A., Grill, R. J. & Tuszynski, M. H. Nerve growth factor-hypersecreting Schwann cell grafts augment and guide spinal cord axonal growth and remyelinate central nervous system axons in a phenotypically appropriate manner that correlates with expression of L1. *J. Comp. Neurol.* **413**, 495–506 (1999).
38. Caroni, P. & Schwab, M. E. Two membrane protein fractions from rat central myelin with inhibitory properties for neurite growth and fibroblast spreading. *J. Cell Biol.* **106**, 1281–1288 (1988).
39. Caroni, P., Savio, T. & Schwab, M. E. Central nervous system regeneration: oligodendrocytes and myelin as non-permissive substrates for neurite growth. *Prog. Brain Res.* **78**, 363–370 (1988).
40. Chen, M. S. et al. Nogo-A is a myelin-associated neurite outgrowth inhibitor and an antigen for monoclonal antibody IN-1. *Nature* **403**, 434–439 (2000).
41. Fawcett, J. W. & Asher, R. A. The glial scar and central nervous system repair. *Brain Res. Bull.* **49**, 377–391 (1999).
42. Fitch, M. T. & Silver, J. Glial cell extracellular matrix: boundaries for axon growth in development and regeneration. *Cell Tissue Res.* **290**, 379–384 (1997).
43. Bernhardt, R. R. Cellular and molecular basis of axonal regeneration in the fish central nervous system. *Exp. Neurol.* **157**, 223–240 (1999).
44. Goodman, C. S. Mechanisms and molecules that control growth cone guidance. *Annu. Rev. Neurosci.* **19**, 341–377 (1996).
45. Tessier-Lavigne, M. Eph receptor tyrosine kinases, axon repulsion, and the development of topographic maps. *Cell* **82**, 345–348 (1995).
46. Pasterkamp, R. J. et al. Expression of the gene encoding the chemorepellent semaphorin III is induced in the fibroblast component of neural scar tissue formed following injuries of adult but not neonatal CNS. *Mol. Cell Neurosci.* **13**, 143–166 (1999).
47. Jaworski, D. M., Kelly, G. M. & Hockfield, S. Intracranial injury acutely induces the expression of the secreted isoform of the CNS-specific hyaluronan-binding protein BEHAB/brevican. *Exp. Neurol.* **157**, 327–337 (1999).
48. Haas, C. A., Rauch, U., Thon, N., Merten, T. & Deller, T. Entorhinal cortex lesion in adult rats induces the expression of the neuronal chondroitin sulfate proteoglycan neurocan in reactive astrocytes. *J. Neurosci.* **19**, 9953–9963 (1999).
49. Lemons, M. L., Howland, D. R. & Anderson, D. K. Chondroitin sulfate proteoglycan immuno-reactivity increases following spinal cord injury and transplantation. *Exp. Neurol.* **160**, 51–65 (1999).

50. Fidler, P. S. *et al.* Comparing astrocytic cell lines that are inhibitory or permissive for axon growth: the major axon-inhibitory proteoglycan is NG2. *J. Neurosci.* **19**, 8778–8788 (1999).
51. Fitch, M. T. & Silver, J. Activated macrophages and the blood-brain barrier: inflammation after CNS injury leads to increases in putative inhibitory molecules. *Exp. Neurol.* **148**, 587–603 (1997).
52. Roitbak, T. & Sykova, E. Diffusion barriers evoked in the rat cortex by reactive astrogliosis. *Glia* **28**, 40–48 (1999).
53. Zuo, J., Neubauer, D., Dyess, K., Ferguson, T. A. & Muir, D. Degradation of chondroitin sulfate proteoglycan enhances the neurite-promoting potential of spinal cord tissue. *Exp. Neurol.* **154**, 654–662 (1998).
54. Bush, T. G. *et al.* Leukocyte infiltration, neuronal degeneration, and neurite outgrowth after ablation of scar-forming, reactive astrocytes in adult transgenic mice. *Neuron* **23**, 297–308 (1999).
55. Stichel, C. C. *et al.* Inhibition of collagen IV deposition promotes regeneration of injured CNS axons. *Eur. J. Neurosci.* **11**, 632–646 (1999).
56. Thallmair, M. *et al.* Neurite growth inhibitors restrict plasticity and functional recovery following corticospinal tract lesions. *Nature Neurosci.* **1**, 124–131 (1998).
57. Keirstead, H. S., Morgan, S. V., Wilby, M. J. & Fawcett, J. W. Enhanced axonal regeneration following combined demyelination plus schwann cell transplantation therapy in the injured adult spinal cord. *Exp. Neurol.* **159**, 225–236 (1999).
58. Huang, D. W., McKerracher, L., Braun, P. E. & David, S. A therapeutic vaccine approach to stimulate axon regeneration in the adult mammalian spinal cord. *Neuron* **24**, 639–647 (1999).
59. Davies, S. J. *et al.* Regeneration of adult axons in white matter tracts of the central nervous system. *Nature* **390**, 680–683 (1997).
60. Pettigrew, D. B. & Crutcher, K. A. White matter of the CNS supports or inhibits neurite outgrowth in vitro depending on geometry. *J. Neurosci.* **19**, 8358–8366 (1999).
61. Bredesen, D. E. Neural apoptosis. *Ann. Neurol.* **38**, 839–851 (1995).
62. Harvey, N. L. & Kumar, S. The role of caspases in apoptosis. *Adv. Biochem. Eng. Biotechnol.* **62**, 107–128 (1998).
63. Blomer, U., Kafri, T., Randolph-Moore, L., Verma, I. M. & Gage, F. H. Bcl-xL protects adult septal cholinergic neurons from axotomized cell death. *Proc. Natl Acad. Sci. USA* **95**, 2603–2608 (1998).
64. Bisby, M. A. & Tetzlaff, W. Changes in cytoskeletal protein synthesis following axon injury and during axon regeneration. *Mol. Neurobiol.* **6**, 107–123 (1992).
65. Skene, J. H. Axonal growth-associated proteins. *Annu. Rev. Neurosci.* **12**, 127–156 (1989).
66. Benowitz, L. I. & Routtenberg, A. GAP-43: an intrinsic determinant of neuronal development and plasticity. *Trends Neurosci.* **20**, 84–91 (1997).
67. Hopker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M. & Holt, C. Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1. *Nature* **401**, 69–73 (1999).
68. Davies, A. M. The Bcl-2 family of proteins, and the regulation of neuronal survival. *Trends Neurosci.* **18**, 355–358 (1995).
69. Chen, D. F., Schneider, G. E., Martinou, J. C. & Tonegawa, S. Bcl-2 promotes regeneration of severed axons in mammalian CNS. *Nature* **385**, 434–439 (1997).
70. Gomez, T. M. & Spitzer, N. C. In vivo regulation of axon extension and pathfinding by growth-cone calcium transients. *Nature* **397**, 350–355 (1999).
71. Qiu, J., Cai, D. & Filbin, M. T. Glial inhibition of nerve regeneration in the mature mammalian CNS. *Glia* **29**, 166–174 (2000).
72. Song, H. J., Ming, G. L. & Poo, M. M. cAMP-induced switching in turning direction of nerve growth cones. *Nature* **388**, 275–279 (1997).
73. Song, H. *et al.* Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides. *Science* **281**, 1515–1518 (1998).
74. Cai, D., Shen, Y., De Bellard, M., Tang, S. & Filbin, M. T. Prior exposure to neurotrophins blocks inhibition of axonal regeneration by MAG and myelin via a cAMP-dependent mechanism. *Neuron* **22**, 89–101 (1999).
75. Tuszyński, M. H. *et al.* Nerve growth factor delivery by gene transfer induces differential outgrowth of sensory, motor, and noradrenergic neurites after adult spinal cord injury. *Exp. Neurol.* **137**, 157–173 (1996).
76. Stichel, C. C., Hermanns, S., Lausberg, F. & Müller, H. W. Effects of schwann cell suspension grafts on axon regeneration in subacute and chronic CNS traumatic injuries. *Glia* **28**, 156–165 (1999).
77. Kawaja, M. D. & Gage, F. H. Reactive astrocytes are substrates for the growth of adult CNS axons in the presence of elevated levels of nerve growth factor. *Neuron* **7**, 1019–1030 (1991).
78. Li, Y., Field, P. M. & Raisman, G. Repair of adult rat corticospinal tract by transplants of olfactory ensheathing cells. *Science* **277**, 2000–2002 (1997).
79. Ramon-Cueto, A., Plant, G. W., Avila, J. & Bunge, M. B. Long-distance axonal regeneration in the transected adult rat spinal cord is promoted by olfactory ensheathing glia transplants. *J. Neurosci.* **18**, 3803–3815 (1998).
80. Ramon-Cueto, A., Cordero, M. I., Santos-Benito, F. F. & Avila, J. Functional recovery of paraplegic rats and motor axon regeneration in their spinal cords by olfactory ensheathing glia. *Neuron* **25**, 425–435 (2000).
81. Xu, X. M., Zhang, S. X., Li, H., Aebischer, P. & Bunge, M. B. Regrowth of axons into the distal spinal cord through a Schwann-cell-seeded mino-channel implanted into hemisectioned adult rat spinal cord. *Eur. J. Neurosci.* **11**, 1723–1740 (1999).
82. Gautier, S. E. *et al.* Poly(alpha-hydroxyacids) for application in the spinal cord: resorbability and biocompatibility with adult rat Schwann cells and spinal cord. *J. Biomed. Mater. Res.* **42**, 642–654 (1998).
83. Houle, J. D. & Ye, J. H. Survival of chronically-injured neurons can be prolonged by treatment with neurotrophic factors. *Neuroscience* **94**, 929–936 (1999).
84. Matyszak, M. K. Inflammation in the CNS: balance between immunological privilege and immune responses. *Prog. Neurobiol.* **56**, 19–35 (1998).
85. Brewer, K. L., Bethea, J. R. & Yezierski, R. P. Neuroprotective effects of interleukin-10 following excitotoxic spinal cord injury. *Exp. Neurol.* **159**, 484–493 (1999).
86. Logan, A., Green, J., Hunter, A., Jackson, R. & Berry, M. Inhibition of glial scarring in the injured rat brain by a recombinant human monoclonal antibody to transforming growth factor- β 2. *Eur. J. Neurosci.* **11**, 2367–2374 (1999).
87. DiProspero, N. A., Meiners, S. & Geller, H. M. Inflammatory cytokines interact to modulate extracellular matrix and astrocytic support of neurite outgrowth. *Exp. Neurol.* **148**, 628–639 (1997).
88. Blinzinger, K. & Kreutzberg, G. Displacement of synaptic terminals from regenerating motoneurons by microglial cells. *Z. Zellforsch. Mikrosk. Anat.* **85**, 145–157 (1968).
89. Streit, W. J. *et al.* Cytokine mRNA profiles in contused spinal cord and axotomized facial nucleus suggest a beneficial role for inflammation and gliosis. *Exp. Neurol.* **152**, 74–87 (1998).
90. Popovich, P. G. *et al.* Depletion of hematogenous macrophages promotes partial hindlimb recovery and neuroanatomical repair after experimental spinal cord injury. *Exp. Neurol.* **158**, 351–365 (1999).

Acknowledgements

We thank E. Brandon, S. Colamarino, M.-C. Senut, L. Shihabuddin, H. van Praag, B. Benish and L. Horky for providing input during the preparation of this review. We appreciate the editorial assistance of M. L. Gage and the assistance of E. Grabowski in the preparation of illustrations. We are grateful for the continued support of The Christopher Reeve Paralysis Foundation, The Lookout Fund, The Parkinson's Disease Foundation and the National Institutes of Health. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Correspondence should be addressed to F.H.G. (e-mail: gage@salk.edu).

Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor

Naoki Kunishima^{*‡}, Yoshimi Shimada[†], Yuji Tsuji[†], Toshihiro Sato[†], Masaki Yamamoto[‡], Takashi Kumasaka[‡], Shigetada Nakanishi[§], Hisato Jingami[†] & Kosuke Morikawa^{*}

Departments of ^{*}Structural Biology and [†]Molecular Biology, Biomolecular Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

[‡]Structural Biophysics Laboratory, RIKEN Harima Institute, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan

[§]Department of Biological Sciences, Kyoto University Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan

The metabotropic glutamate receptors (mGluRs) are key receptors in the modulation of excitatory synaptic transmission in the central nervous system. Here we have determined three different crystal structures of the extracellular ligand-binding region of mGluR1—in a complex with glutamate and in two unliganded forms. They all showed disulphide-linked homodimers, whose ‘active’ and ‘resting’ conformations are modulated through the dimeric interface by a packed α -helical structure. The bi-lobed protomer architectures flexibly change their domain arrangements to form an ‘open’ or ‘closed’ conformation. The structures imply that glutamate binding stabilizes both the ‘active’ dimer and the ‘closed’ protomer in dynamic equilibrium. Movements of the four domains in the dimer are likely to affect the separation of the transmembrane and intracellular regions, and thereby activate the receptor. This scheme in the initial receptor activation could be applied generally to G-protein-coupled neurotransmitter receptors that possess extracellular ligand-binding sites.

Glutamate is a principal excitatory neurotransmitter in the mammalian central nervous system and functions in long-term potentiation/depression, learning and memory. Glutamate receptors are membrane proteins, which mediate excitatory transmission on the cellular surface through initial binding of glutamate^{1,2}, and are categorized into ionotropic glutamate receptors (iGluRs) and metabotropic glutamate receptors (mGluRs). The mGluRs are coupled to G proteins and are thus distinct from the iGluRs which internally contain ligand-gated ion-channels. From a structural viewpoint, mGluRs belong to the seven-transmembrane protein family mostly shared by G-protein-coupled receptors, such as rhodopsin and the adrenergic receptor.

The eight subtypes of mGluRs are classified into three subgroups according to sequence similarity, agonist selectivity and effector system differences³: subgroup I (mGluR1 and -5), subgroup II (mGluR2 and -3) and subgroup III (mGluR4, -6, -7 and -8). They share sequence similarity with the calcium-sensing receptor⁴ and the pheromone receptor^{5,6} to form the mGluR family. The mGluR structure is divided into three regions: the extracellular region, the seven-spanning transmembrane region and the cytoplasmic region (Fig. 1a). The extracellular region is further divided into the ligand-binding region (LBR) and the cysteine-rich region. LBR has sequence similarity to the leucine/isoleucine/valine-binding protein (LIVBP), which is a bacterial periplasmic binding protein⁷ (PBP), as well as to the extracellular regions of both iGluR⁷ and the γ -aminobutyric acid (GABA)_B receptor⁸. Both chimaeric receptor analysis⁹ and homology model building^{7,10} based on the crystal structure of LIVBP¹¹ have suggested that LBR is responsible for ligand binding. More directly, the extracellular regions of mGluR have been expressed in soluble forms, and have been shown to serve as ligand-binding sites and thus confer ligand-binding selectivity^{12–14}. The mGluR family therefore differs in the ligand-binding mode from the classical rhodopsin-type G-protein-coupled receptors, which bind ligands in pockets in their transmembrane helices. To gain insight into the mechanism of how ligand binding induces a conformational change that is transmitted across the transmembrane region and into the intracellular region in the mGluR family

proteins, we crystallized the LBR of mGluR1 (m1-LBR) and analysed it by X-ray crystallography. Here we report three crystal structures of m1-LBR with and without glutamate, and discuss their structural differences to deduce the implications of receptor activation.

General description of crystal structures

We determined the crystal structures of m1-LBR in two unliganded forms, called free forms I and II, and in a complex form with glutamate (Table 1). The asymmetric unit of all crystal forms contains two molecules of m1-LBR. These two protomers form a homodimer related by a non-crystallographic pseudo-twofold axis (Fig. 1b, c). The global conformations of the dimers substantially differ between the complex form and the free form I, whereas the free form II and the complex form share almost identical conformations, except for ligand binding.

The two protomers are covalently connected by an interprotomer disulphide bridge between Cys 140 of each monomer, which lie in a long disordered segment between the B helix and the D strand. This disulphide bridge has been confirmed by biochemical analyses of the protein with an alanine substitution at Cys 140 (ref. 15). Obviously, this disulphide bond cannot act as a structural scaffold because of its location within the disordered segment. Instead, it most probably functions as an interprotomer linker to increase the effective concentration of a dimeric form of mGluR1 on the cellular surface. Similar reports showing the formation of disulphide-linked dimers of mGluR5 (ref. 16) and the calcium receptor¹⁴ further support the critical role of this residue, in addition to the conservation of the corresponding cysteines in the mGluR family (Fig. 2).

Each protomer consists of the amino-terminal and the carboxy-terminal domains, designated LB1 and LB2, respectively. Only the LB1 domain provides the dimer interfaces in all three crystal forms, and hence the LB1–LB2 relative orientations are independent of the dimer interface construction. These two domains, both with α/β topology, are connected by three short loops on one side of the molecule to form a ‘clamshell’-like shape similar to that found in iGluR¹⁷ or PBPs, such as LIVBP¹¹, lysine/arginine/ornithine-binding

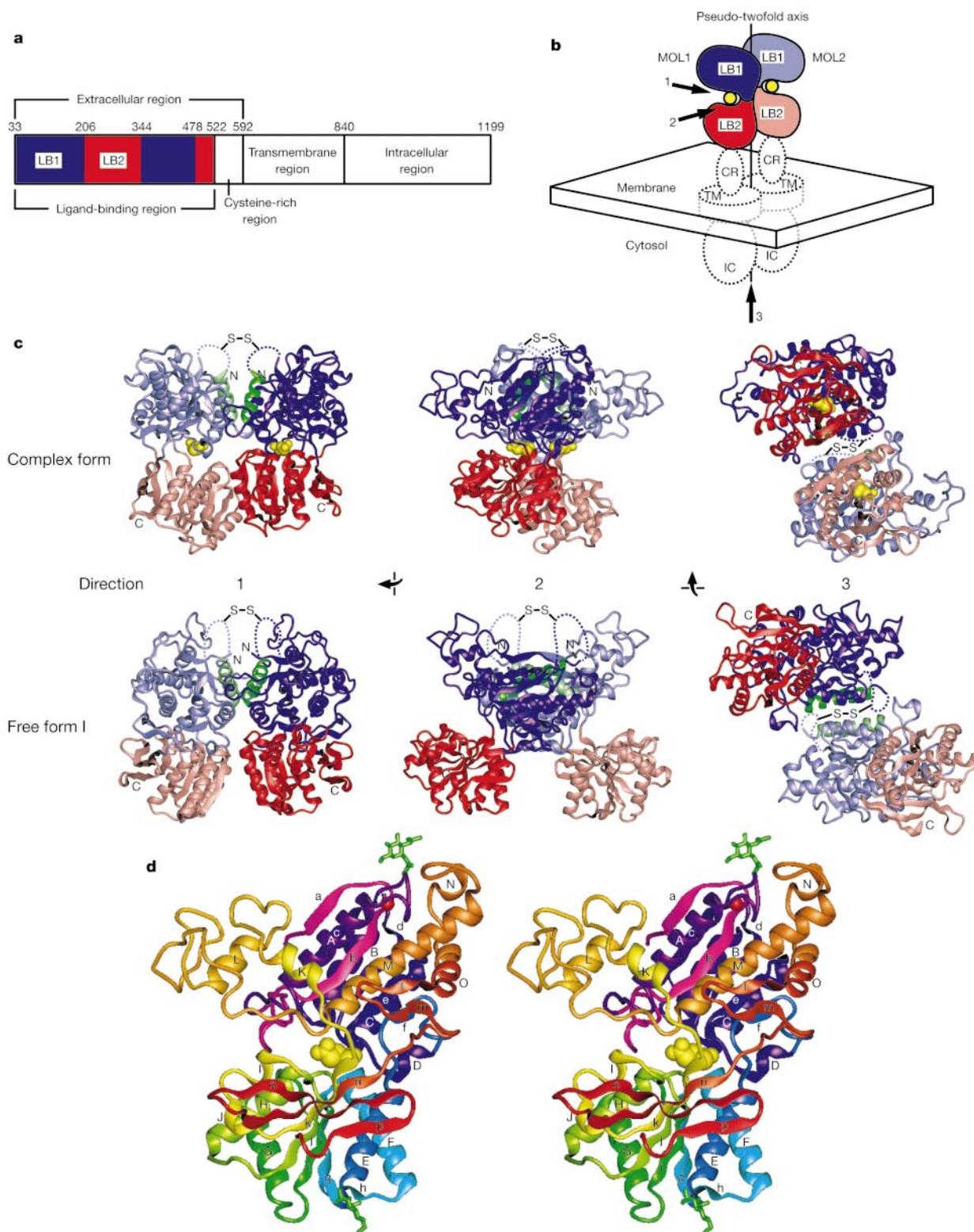


Figure 1 Crystal structures of mGluR1 in the liganded and unliganded states. **a**, Diagram of full-length mGluR1. Functional regions are boxed. The LB1 and LB2 domains, which constitute a ligand-binding region, are coloured blue and red, respectively. Numerical positions of amino-acid residues are indicated according to the primary amino-acid sequence⁴⁰. Protein regions not determined in this study are indicated by open boxes. **b**, Drawing of the spatial arrangements of mGluR1 domains. The MOL1 and MOL2 molecules of the m1-LBR dimer are distinguished by dark and light colouring, respectively. The ligand, glutamate, is shown as yellow spheres. CR, cysteine-rich region; IC, intracellular region; TM, transmembrane region. **c**, Three orthogonal views of the

dimer structures of the complex form and the free form I. The perspectives are indicated by arrows in **b**. Bound glutamate molecules are shown as yellow space-filling models. Disordered regions with a potential interprotomer disulphide bridge are indicated by dotted lines. The B helix, constituting the dimer interface between subunits, is coloured green. **d**, Stereoview of MOL1 of the complex form. The ribbon model, viewed from direction 2 in **c**, is coloured in a rainbow gradation from purple (N terminus) to red (C terminus). Helices and strands are indicated by upper and lower cases, respectively. Sugar moieties and a bound metal ion are depicted by green licorice models and a red sphere, respectively.

protein¹⁸ (LAOBP), polyamine-binding protein^{19,20} and glutamine-binding protein²¹ (GlnBP). As predicted from sequence analysis⁷, the main-chain folding of m1-LBR is similar to that of LIVBP; the r.m.s. deviations for the corresponding C α atoms are about 1.8 Å for each domain.

Glutamate is bound in an interdomain crevice, as found in iGluR or PBPs. In common with these proteins, m1-LBR shows an open-closed conformation characterized by the different spatial arrangements between the LB1 and LB2 domains. We designate the relatively closed subunit of the m1-LBR dimer as MOL1 and the other as MOL2. The most characteristic feature of m1-LBR is the strikingly different dimer configurations, which are dependent on the dimer interface construction.

We found a bound metal ion in the loop region between the A helix and the C strand (Figs 1d, 2). From its hexavalent coordination and temperature factor, this metal is most likely to be Mg²⁺, although Ca²⁺ cannot be ruled out. mGluR1 has been reported to bind di-/trivalent cations near the glutamate-binding site and function as a cation receptor²². However, the found metal-binding position appears to be far from the reported cation-binding site. As

we added neither magnesium nor calcium during the purification and crystallization, the metal appears to bind to m1-LBR more tightly than the reported affinity²². Therefore, it may be a scaffold factor for the LB1 domain, rather than a second physiological ligand.

The B helix adjacent to the N terminus of the long disordered segment shows a notable conformational change between the free form I and the complex form. The residues 125–131 are disordered in the complex form, whereas in the free form I they are incorporated into the B helix to form a helical extension of about two turns on the C-terminal side (Fig. 2). The extended helical residues in the free form I interact with those in the counterpart of the m1-LBR dimer. The same residues in the free form II are disordered, like those in the complex form. These findings suggest that the elongation of the B helix is more directly coupled with the domain arrangements rather than with glutamate binding itself.

Ligand recognition

In the complex form, both MOL1 and MOL2 bind glutamate at

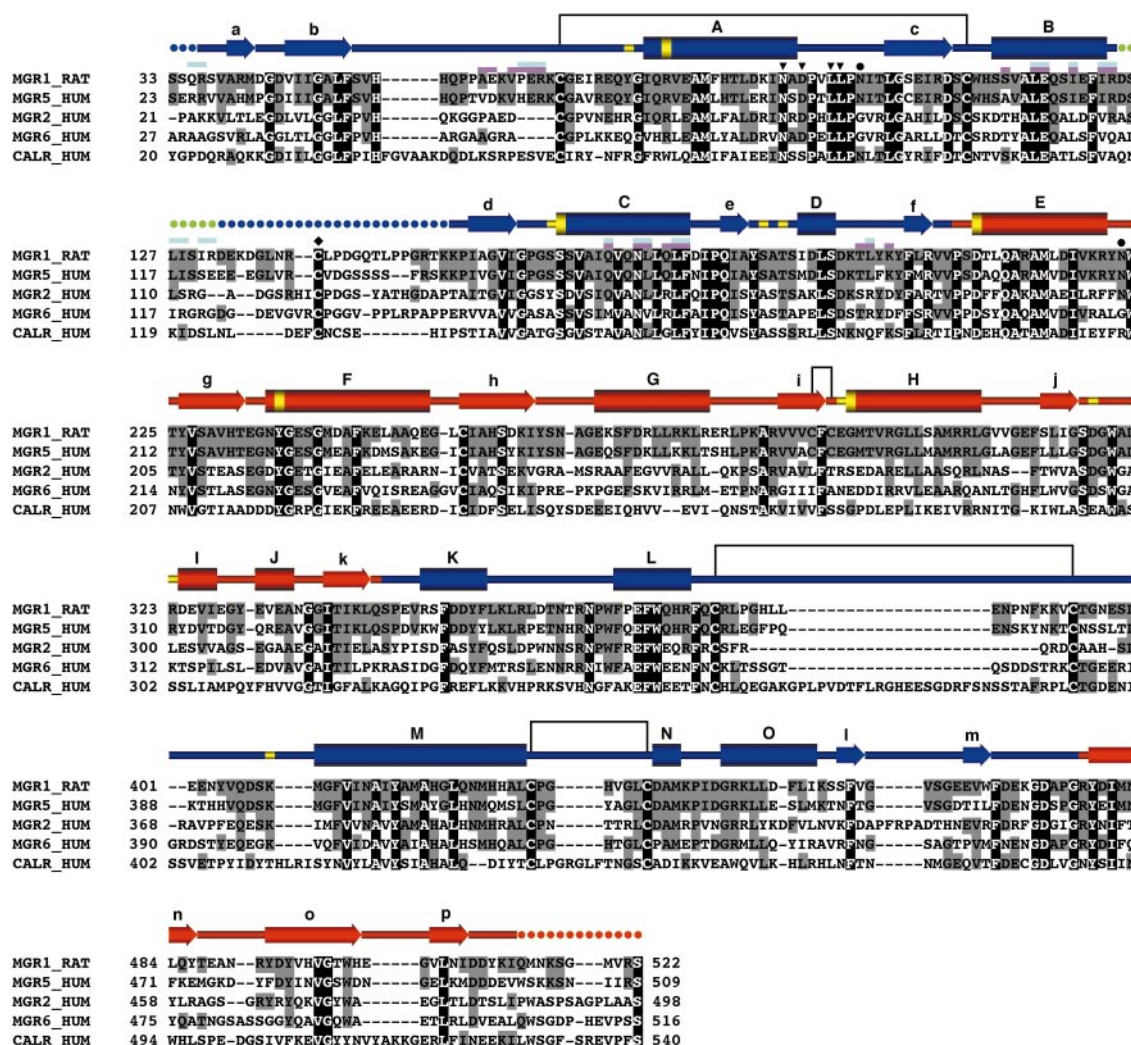


Figure 2 Secondary structure assignment (PROCHECK³⁷) and multiple sequence alignments. Secondary structure in MOL1 of the complex form is displayed over the sequences: helices are shown as cylinders; strands as arrows, with the same labelling as in Fig. 1d; disordered residues are shown as dots. The LB1 and LB2 domains, the ligand-binding sites, and the additionally visible region in the free form I are coloured blue, red, yellow and green, respectively. Intraprotomer disulphide bridges are indicated by thin solid lines connecting cysteines. Residues of interest are highlighted beneath the

secondary structure: cyan and magenta bars indicate residues involved in the dimer interfaces of the free form I and the complex form, respectively; arrowheads and filled circles indicate residues coordinating with the metal ion and glycosylated residues, respectively; filled diamond indicates the cysteine residue forming the interprotomer disulphide bridge. The sequences of selected mGluR family proteins, mGluR1, -2, -5 and -6 (refs 40–43) and the calcium receptor⁴⁴, are aligned together and shown with the shading highlighting conserved sequences.

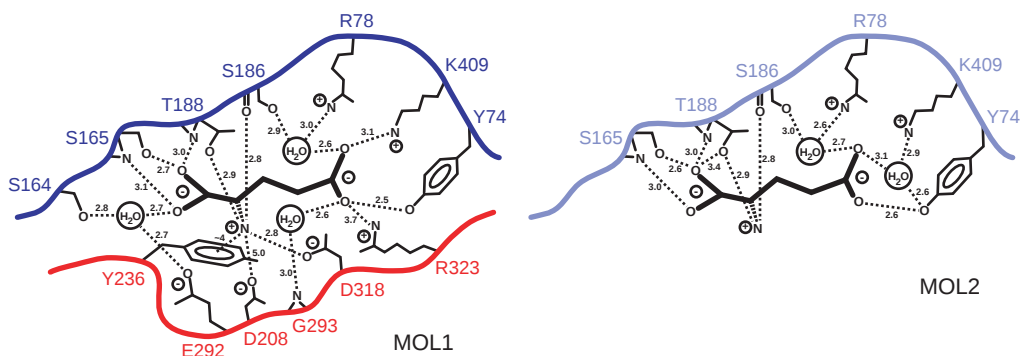


Figure 3 Representation of glutamate binding. Polar interactions of glutamate with MOL1 and MOL2 are denoted by dotted lines with distances. Domain colouring is as in Fig. 1b.

similar sites (Fig. 3). A large cleft formed by the bi-lobed architecture involves two interfaces that contact the glutamate: one lies in the LB1 domain (LB1 interface) and the other in the LB2 domain (LB2 interface). The structures of the LB1 interface are essentially identical between MOL1 and MOL2, except for the positions of a few bound water molecules. By contrast, the LB2 interface with the open conformation is not used for glutamate binding by MOL2. The similar averaged temperature factors of the two bound glutamates indicate essentially identical mobility between MOL1 and MOL2, suggesting a predominant role for the LB1 domain in anchoring the ligand.

The LB1 interfaces with a buried surface area of $168 \text{ \AA}^2$ are dominated by direct hydrogen bonds, while water-mediated hydrogen bonds and salt bridges are predominant in the LB2 interface

with a $135 \text{ \AA}^2$ buried surface area. A cation– π interaction²³ between glutamate and Tyr 236 may contribute additionally to stabilizing the complex.

All of the potential ligand-binding residues identified by mutational analyses^{7,24} are hydrogen bonded to glutamate in the LB1 interfaces. Our recent mutational analysis has confirmed that the hydrogen bond of Tyr 74 with glutamate is important for ligand binding (data not shown). Seven polar residues, which contribute to the ligand recognition, are conserved among all of the mGluR subtypes (Fig. 2). On the other hand, five polar residues are conserved only in subgroup I (mGluR1 and -5). These five residues may be related to the ligand preference of subgroup I. The two glutamate-binding schemes in the closed and open bi-lobed structure provide a structural basis for drug design.

Table 1 Statistics from the crystallographic analysis

Native data statistics*								
Data set	Space group	Unit-cell dimensions		Resolution (Å)	Completeness (%)	Multiplicity	R_{merge} (%)	$I/\sigma(I)$
		(Å)	(deg)					
Complex form	$P2_1$	$a = 83.4$ $b = 95.2$ $c = 97.5$	$\beta = 114.9$	2.2 (2.25–2.20)	98.9 (96.2)	3.26 (2.83)	4.5 (13.7)	17.5 (8.6)
Free form I	$P4_12_12$	$a = 111.4$ $c = 293.7$		3.7 (3.83–3.70)	99.9 (100.0)	6.08 (5.96)	14.6 (52.0)	11.6 (3.5)
Free form II	$P2_1$	$a = 84.8$ $b = 94.5$ $c = 95.5$	$\beta = 113.2$	4.0 (4.14–4.00)	89.1 (77.3)	2.55 (2.51)	18.1 (25.0)	5.4 (3.6)
MAD phasing from the K_2PtCl_4 derivative of the complex form (17.0–4.0 Å)†								
Data set	Wavelength (Å)	Completeness (%)	R_{merge} (%)	Number of sites	Phasing power (acentric/centric)			Mean FOM
					Isomorphous	Anomalous		
Edge	1.0720	99.4	4.3	6	3.25/2.48	1.65/-		0.442
Peak	1.0717	99.6	4.8	–	3.12/2.47	1.87/-		–
Remote	1.0200	99.5	4.6	–	–	1.82/-		–
Refinement statistics‡								
	Complex form		Free form I		Free form II			
Resolution range (Å)	17.0–2.2		12.0–3.7		20.0–4.0			
R factor (%)	19.6 (65,274 reflections)		24.4 (18,693 reflections)		25.4 (10,016 reflections)			
R_{free} (%)	22.7 (3,497 reflections)		28.7 (950 reflections)		32.8 (534 reflections)			
Ramachandran (%)								
Favoured	89.1		89.3		85.4			
Generous	0.4		0.5		0.4			
Disallowed	0.0		0.0		0.0			
R.m.s. deviations from ideality								
Bond lengths (Å)	0.006		0.013		0.014			
Bond angles (deg)	1.2		1.7		1.8			
B factor (Å ²)	0.80 (1.14)		1.48 (1.56)		1.08 (1.16)			

* Values in parentheses indicate statistics for the highest resolution shells. $R_{\text{merge}} = \sum_i \sum_j |I_j - \langle I \rangle| / \sum_i \sum_j I_j$, where $\langle I \rangle$ is the mean intensity of i th unique reflection, and I_j is the intensity of j th observation.
† Phasing power = $(F(H)/E)$, where $F(H)$ is the r.m.s. amplitude of the heavy atom structure factor and E is the residual lack of closure error. Mean figure of merit (FOM) = $(\sum P(\alpha) \exp(i\alpha) / \sum P(\alpha))$, where $P(\alpha)$ is the phase probability at the angle α .
‡ The residues in disordered regions were eliminated from the model: S33–Q35 (MOL1), S33–S34 (MOL2), D125–K153 and Q513–S522 of the complex form; S33–S34, D132–K153 and Q513–S522 of the free form I; S33–Q35, D125–K153 and Q513–S522 of the free form II. R factor = $\sum_i |F(\text{obs}) - F(\text{calc})| / \sum_i F(\text{obs})$, where $F(\text{obs})$ and $F(\text{calc})$ are the observed and calculated native structure factors, respectively. $R_{\text{free}} = R$ factor calculated using 5% of total reflections, which were chosen randomly and omitted from the refinement. All observed reflections, except for those for the calculation of R_{free} , were used for the refinement. R.m.s. deviations of B factors were calculated for main-chain atoms and side-chain atoms (in parentheses).

Interdomain movement

The open–closed conformational change of m1-LBR can be described as a rotation about an axis roughly passing across the three connecting loops (Fig. 1c, d). The difference in LB1–LB2 interdomain orientations between MOL1 (closed conformer) and MOL2 (open conformer) in the complex form is 31° , which is smaller than the 52° in LAOBP¹⁸ and 56° in GlnBP²¹ (see Methods). We compared the determinants for interdomain movements, the ϕ and ψ backbone dihedral angles on the three connecting loops, between the two conformers. The largest changes were 47° in the ϕ angle at Gly 477 and 61° in the ψ angle at Arg 478 on the same loop. These values are comparable to the 52° change that occurs in the ψ angle at Ala 90 in LAOBP¹⁸, whereas the other two loops show much smaller angle differences spread over many residues.

In contrast to the closed and open conformers in the complex form, the free form I contains two open conformers, similar to each other in the overall shapes but different in detail (Fig. 1c). The two closed conformers and the four open conformers, available in all three different crystal forms, allowed us to estimate the robustness of both conformers by comparing the interdomain orientations. The two closed conformers are identically orientated within an azimuth angle of 1° , whereas the four open conformers show a polymorphism of about 10° . This orientational restraint of the closed conformers seems to be derived from the direct contact between the LB1 and LB2 domains, which is shared by the surroundings of the ligand interfaces in the free form II and the complex form. This contact with a buried surface area of $477 \text{ \AA}^2$ involves six polar and many non-polar interactions.

The free form II maintains the same dimer conformations as those of the complex form, even in the absence of ligand. This suggests that the interdomain orientation of the ligand-free m1-LBR exchanges in solution between the open and closed conformations. From a comparison of the unliganded and complex forms of the LAOBP crystal structures, it was proposed that the unliganded protein undergoes a dynamic change between an open and a closed conformation, and that the role of the ligand is to stabilize the closed conformation¹⁸. Furthermore, as reported for LAOBP, we could not find any evidence for a direct link between the ligand binding and the conformational change of the three connecting loops in m1-LBR. Together, we conclude that one role of glutamate may be to stabilize the closed conformation of the bi-lobed m1-LBR protomer in dynamic equilibrium.

Relocation of dimer interfaces

On the dimer interfaces, the two LB1 domains rotate by about 70° between the free form I and the complex form (Fig. 4a). The rotation axis is approximately perpendicular to these interfaces and passes through the vicinity of Ile 120 and Leu 174 (Fig. 4b, c). This rotation, which is uncoupled from the intraprotomer open–closed conformational change, solely determines the interprotomer orientations of the non-crystallographic symmetry (NCS) related LB1 domains. We designated the two distinct interprotomer orientations of the LB1 domains in the free form I and the complex form as the ‘R’ and ‘A’ conformations, respectively. A comparison between the free form II and the complex form revealed that the A conformation is fixed within an azimuth angle of 3° . A restraint comparable to this angle would be applied to the R conformation, because the buried surface areas of the two interfaces are similar: $906 \text{ \AA}^2$ for the R conformation and $793 \text{ \AA}^2$ for the A conformation. The central hydrophobic cores of the two distinct interfaces comprise residues Leu 116, Ile 120, Leu 127, Leu 174 and Leu 177, which are well conserved in the mGluR family (Figs 2, 4c).

The dimer interface related by NCS mainly consists of the B and C helices, including the extension of the B helix in the free form I (Fig. 4b). This helical extension appears to predominantly differentiate between the R and A conformations. The angles between the B and C helices of each LB1 domain are estimated to be 140° in the R

conformer and 70° in the A conformer. The mode of helical packing in this interface appears to be different from that of the globin fold or the four-helix bundle²⁵, where the crossing angles are usually 20° or 50° . The packing in the dimer interface appears to be looser than that in the internal protein folding. This looseness may be favourable for the smooth conversion between the R and A conformations.

Multiple conformations of the LBR dimer

The entire conformations of the m1-LBR dimer in the three crystal forms can be defined as a combination of the intra and inter-protomer conformations: open–open/R for the free form I and closed–open/A for both of the free form II and the complex form (Fig. 5). The latter, closed–open/A conformation is assumed to be

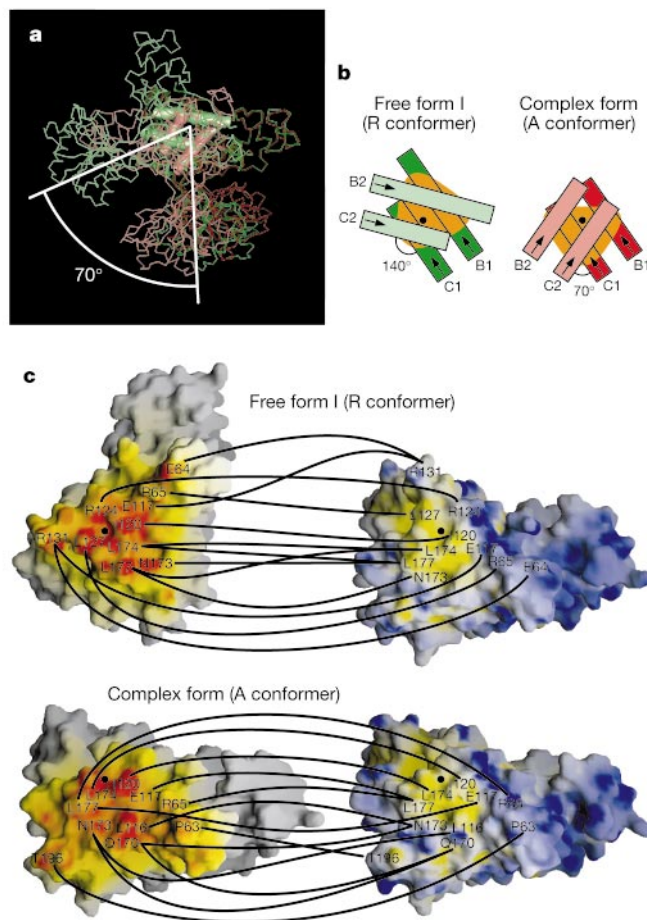


Figure 4 Relocation of dimer interfaces. **a**, Superposition of the free form I (green) and the complex form (red) after the alignment of the corresponding Cα atoms in the LB1 domains of MOL1. MOL1 and MOL2 are distinguished by dark and light colouring, respectively. The two molecules in the different states are viewed along the rotation axis of the dimer interfaces. The helices B and C, which mainly contribute to the dimeric interactions, are represented as ribbon diagrams in MOL2 with the same colouring as that of the subunits. **b**, Close-up view of the dimer interfaces. Perspective and colouring of the receptors are the same as those in **a**. Helices B and C are represented as rectangles with arrows that denote sequence directions. Dimer interfaces and their rotation axes are depicted as orange planes and filled circles, respectively. **c**, Open-book view of dimer interfaces. GRASP45 models of the LB1 domains, viewed from the right side of **a**, are disassembled in the manner of an ‘open book’ to reveal the dimer interfaces. Hydrophobicity based on the reported hydropathy values⁴⁶ and buried surface area were mapped, respectively, onto the right MOL1 models and the left MOL2 models by gradation colouring. The blue and the yellow in MOL1 represent the hydrophilic and hydrophobic regions, respectively; the yellow and the red in MOL2 indicate the low and high values of buried surface area, respectively. Residue–residue interactions and rotation axes of the dimer interfaces are indicated.

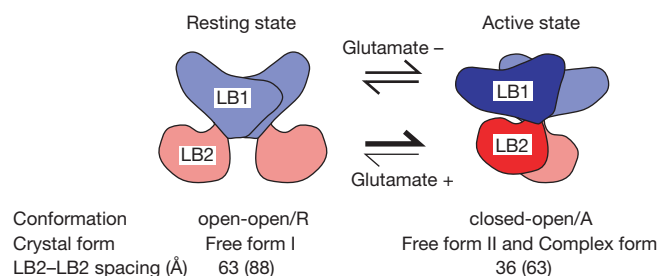


Figure 5 Diagram of the two states of the m1-LBR in dynamic equilibrium. Open and closed conformations of m1-LBR are distinguished by light and dark colouring, respectively. LB2-LB2 spacings are shown between the centres of masses or C termini (in parentheses) of the two LB2 domains in each conformer.

trapped by crystal packing, because of the lack of connection between the asymmetric protomer combination and the interprotomer A conformation. This implies that unidentified conformations of m1-LBR with various combinations might be present in solution. We confirmed the feasibility of other putative conformers by model building (see Methods). This examination allowed us to construct four additional conformers of the m1-LBR dimers without steric clashes: closed-open/R, closed-closed/R, open-open/A and closed-closed/A.

If glutamate binding changes the equilibrium of such multi-conformers and thereby activates the receptor, then the spatial arrangements of the two LB2 domains near the cellular membrane would be crucial for signal transduction (Figs 1b, 5). Thus, we estimated the spacings between the two LB2 domains in each of the six conformers (see Fig. 5 legend). Notably, in spite of the large variety of LB2-LB2 orientations, there are only two spacings with a roughly 26 Å difference determined exclusively by the interprotomer R-A conformations (Fig. 5). Notably, the crystal structures indicate that glutamate binding settles the A conformation, which should be defined as the 'active' state, and hence the R conformation should correspond to the 'resting' state. This notion implies that glutamate binding should be the determinant for the R-A conformational change, perhaps through a subtle structural change that occurs on binding. The limited resolutions of the ligand-free crystals prevent manifesting a clear structural basis for the stabilization mechanism, although a loop region (residues 59–67), including a conserved disulphide bridge, appears to have an important role, because of its proximity to both the glutamate-binding site and the extending region of the B helix.

Implications of receptor activation

The initial activation mode of mGluR1 is totally different from those postulated in the classical rhodopsin-type G-protein-coupled receptors including rhodopsin²⁶ and the adrenergic receptor²⁷. It is intriguing to compare the activation mechanism of mGluR1 with cytokine-receptor recognition. Structural studies revealed that cytokine binding causes the assembly and/or the relocation of their extracellular domains of receptor molecules^{28,29}. In activation of the dimeric erythropoietin receptor (EPOR), the conformational rearrangement of the extracellular region coupled with ligand binding induces the access of the transmembrane and cytoplasmic domains^{30,31}. Similarly, the interdomain and interprotomer rearrangements of the m1-LBR dimer may cause the approach of two transmembrane regions linked to intracellular regions (Fig. 5), although we have to consider the cysteine-rich region with an unknown 3D structure (Fig. 1b). Alternatively, glutamate with a much smaller size and lower affinity does not use the induced-fit activation mechanism, and the ligand binding increases only the population of activated forms by stabilizing them among the several conformations that are in dynamic equilibrium.

The crystal structures of m1-LBR also suggest roles for the LBR homologous regions (LBR-HRs) of related receptors, such as the GABA_B receptor and iGluR. The extracellular LBR-HRs of both receptors possess a short stretch of hydrophobic residues, corresponding to the C helix of the m1-LBR dimer interface, suggesting the conservation of the oligomerization interface among the three receptors. In the case of the GABA_B receptor, a heterodimerization of two distinct subunits with homologous sequences is essential for receptor expression and activation^{32–34}. Thus, a large variety of interdomain and intersubunit conformations of LBR-HR might be important for the GABA_B receptor activation. Similarly, in the case of iGluR, which forms a multisubunit ion channel, LBR-HR at the N terminus of the receptor might regulate the function of the ion channel by using an additional extracellular interface for oligomerization.

In conclusion, we have shown that the dimeric ligand-binding region of mGluR adopts multiple conformations, where the activated domain arrangements of the dimer are in equilibrium with other states and are selectively stabilized by glutamate binding. Analogous multiple conformations in a multisubunit receptor, formed as a result of the relocations of both the interdomain and intersubunit interfaces, may be one of the fundamental principles of the activation mechanism of receptors that bind small ligands in extracellular regions. □

Methods

Crystallization and data collection

We expressed and purified the LBR of rat mGluR1 (residues 33–522) as described¹⁵. The complex form grew in hanging drops containing 20% PEG4000, 1 mM L-glutamate, 10 mg ml⁻¹ m1-LBR and 200 mM HEPES-NaOH pH 7.5. Heavy atom derivatives were prepared by soaking crystals in 5 mM K₂PtCl₄ for 5 h or by expressing m1-LBR in a medium containing selenomethionine. The free form I was grown in 1.9 M ammonium sulphate solution, containing 10 mg ml⁻¹ m1-LBR and 200 mM Tris-HCl pH 8.5, and the free form II in a 20% PEG4000 solution, containing 10 mg ml⁻¹ m1-LBR, 100 mM ammonium sulphate and 100 mM Tris-HCl pH 7.5. We collected diffraction data with cryocooling at 100 K on undulator beamlines BL45XU (complex form) and BL24XU (free forms) at SPring-8, Hyogo, Japan. Cryosolvents were artificial mother liquor containing 20–30% glycerol. We processed all data with DENZO/SCALEPACK³⁵.

Structure determination

We used multiwavelength anomalous dispersion (MAD) for the phase determination of the platinum derivative of the complex form. MAD phasing from six platinum sites was performed with the program SHARP³⁶ at 4.0 Å resolution. Phase extension to higher resolution data was unsuccessful; however, an initial atomic model containing most of the secondary structure in m1-LBR could be built on the 4.0 Å resolution map, after phase improvement using the program DM³⁷. In this process, we found 22 selenium sites in a difference Fourier map from the selenomethionine derivative. These selenium sites and the atomic coordinates of LIVBP (PDB code 2LIV) were effectively used as guides in the model building. Further phase improvement by molecular averaging and by phase combination with the partial model phases allowed us to build a model for the entire complex molecule. This model was refined with the program CNS³⁸ followed by manual rebuilding. We applied bulk-solvent correction, NCS restraints and anisotropic B factor corrections. The NCS restraints were released in the final two rounds of the refinement. During the refinement, a few bound HEPES molecules were found in regions sufficiently far from both the glutamate-binding site and the dimer interface. The final model contains 897 residues of a m1-LBR dimer, 2 L-glutamate molecules, 642 water molecules, 4 N-acetyl-glucosamine residues, 3 HEPES molecules and 2 metal ions. The model of Cys 289 had two conformers probably produced by radiation damage during data collection, as reported for acetylcholine esterase and lysozyme crystal structures³⁹.

The molecular replacement analysis with AMORE³⁷, using modified coordinates of the complex form as a search model, allowed the structure determination of the free form I crystal. The combination of molecular replacement phases with single isomorphous replacement phases from a Pt derivative provided the identical conformation. We refined the model using CNS with bulk-solvent correction, tight NCS restraints (300 kcal mol⁻¹ Å⁻²) except for the regions of crystal contact, and anisotropic B-factor corrections. In the free form I, bound sulphate ions were found in the moieties essentially identical to the HEPES sites in the complex form. The final model contains 912 amino-acid residues of an m1-LBR dimer, an N-acetyl-glucosamine residue and 4 sulphate ions. The structure of the free form II was solved by rigid-body refinement with CNS, using the coordinates of the complex form. For the calculation, we regarded the four domains in the m1-LBR dimer as four independent rigid bodies. Bulk-solvent correction and anisotropic B-factor corrections were applied. Finally, the model was refined with tight restraints. The current model contains 896 residues of an m1-LBR dimer. No obvious electron densities were observed around the glutamate-binding site of the free form II, after simulated annealing using CNS. $F_{\text{obs}} - F_{\text{calc}}$ simulated annealed omit maps for the free forms I and II

were calculated at 3.7 Å and 4.0 Å resolution, respectively, and both of them clearly revealed one by one all α -helices from A to O, confirming that the refined conformations are accurate. A stereochemical analysis of the refined model by PROCHECK³⁷ showed excellent geometry for all crystal forms (Table 1).

We calculated the orientation differences between two identical components and buried surface areas with the programs LSQKAB³⁷ and QUANTA98 (Molecular Simulations), respectively, as described²⁸. Figures 1c, 1d and 4a were drawn in QUANTA98.

Received 1 June; accepted 30 August 2000.

- Nakanishi, S. & Masu, M. Molecular diversity and functions of glutamate receptors. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 319–348 (1994).
- Hollmann, M. & Heinemann, S. Cloned glutamate receptors. *Annu. Rev. Neurosci.* **17**, 31–108 (1994).
- Nakanishi, S. Molecular diversity of glutamate receptors and implications for brain function. *Science* **258**, 597–603 (1992).
- Brown, E. M. *et al.* Cloning and characterization of an extracellular Ca^{2+} -sensing receptor from bovine parathyroid. *Nature* **366**, 575–580 (1993).
- Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–773 (1997).
- Matsunami, H. & Buck, L. B. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–784 (1997).
- O'Hara, P. J. *et al.* The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **11**, 41–52 (1993).
- Kaupmann, K. *et al.* Expression cloning of GABA_B receptors uncovers similarity to metabotropic glutamate receptors. *Nature* **386**, 239–246 (1997).
- Takahashi, K., Tsuchida, K., Tanabe, Y., Masu, M. & Nakanishi, S. Role of the large extracellular domain of metabotropic glutamate receptors in agonist selectivity determination. *J. Biol. Chem.* **268**, 19341–19345 (1993).
- Costantino, G., Macchiarulo, A. & Pellicciari, R. Modeling of amino-terminal domains of group I metabotropic glutamate receptors: structural motifs affecting ligand selectivity. *J. Med. Chem.* **42**, 5390–5401 (1999).
- Sack, J. S., Saper, M. A. & Quiocho, F. A. Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. *J. Mol. Biol.* **206**, 171–191 (1989).
- Okamoto, T. *et al.* Expression and purification of the extracellular ligand binding region of metabotropic glutamate receptor subtype 1. *J. Biol. Chem.* **273**, 13089–13096 (1998).
- Han, G. -M. & Hampson, D. R. Ligand binding to the amino-terminal domain of the mGluR4 subtype of metabotropic glutamate receptor. *J. Biol. Chem.* **274**, 10008–10013 (1999).
- Goldsmith, P. K. *et al.* Expression, purification, and biochemical characterization of the amino-terminal extracellular domain of the human calcium receptor. *J. Biol. Chem.* **274**, 11303–11309 (1999).
- Tsuji, Y. *et al.* Cryptic dimer interface and domain organization of the extracellular region of metabotropic glutamate receptor subtype 1. *J. Biol. Chem.* **275**, 28144–28151 (2000).
- Romano, C., Yang, W.-L. & O'Malley, K. L. Metabotropic glutamate receptor 5 is a disulphide-linked dimer. *J. Biol. Chem.* **271**, 28612–28616 (1996).
- Armstrong, N., Sun, Y., Chen, G.-Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
- Oh, B.-H. *et al.* Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J. Biol. Chem.* **268**, 11348–11355 (1993).
- Sugiyama, S. *et al.* Crystal structure of PotD, the primary receptor of the polyamine transport system in *Escherichia coli*. *J. Biol. Chem.* **271**, 9519–9525 (1996).
- Vassilyev, D. G., Tomitori, H., Kashiwagi, K., Morikawa, K. & Igarashi, K. Crystal structure and mutational analysis of the *Escherichia coli* putrescine receptor. *J. Biol. Chem.* **273**, 17604–17609 (1998).
- Sun, Y.-J., Rose, J., Wang, B.-C. & Hsiao, C.-D. The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* **278**, 219–229 (1998).
- Kubo, Y., Miyashita, T. & Murata, Y. Structural basis for a Ca^{2+} -sensing function of the metabotropic glutamate receptors. *Science* **279**, 1722–1725 (1998).
- Gallivan, J. P. & Dougherty, D. A. Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **96**, 9459–9464 (1999).
- Hampson, D. R. *et al.* Probing the ligand-binding domain of the mGluR4 subtype of metabotropic glutamate receptor. *J. Biol. Chem.* **274**, 33488–33495 (1999).
- Chothia, C., Levitt, M. & Richardson, D. Structure of proteins: packing of α -helices and pleated sheets. *Proc. Natl. Acad. Sci. USA* **74**, 4130–4134 (1977).
- Khorana, H. G. Rhodopsin, photoreceptor of the rod cell: an emerging pattern for structure and function. *J. Biol. Chem.* **267**, 1–4 (1992).
- Strader, C. D., Fong, T.-M., Tota, M. R., Underwood, D. & Dixon, R. A. F. Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* **63**, 101–132, (1994).
- Aritomi, M. *et al.* Atomic structure of the GCSF-receptor complex showing a new cytokine-receptor recognition scheme. *Nature* **401**, 713–717 (1999).
- Deller, M. C. & Jones, E. Y. Cell surface receptors. *Curr. Opin. Struct. Biol.* **10**, 213–219 (2000).
- Livnah, O. *et al.* Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987–990 (1999).
- Remy, I., Wilson, I. A. & Minchick, S. W. Erythropoietin receptor activation by a ligand-induced conformation change. *Science* **283**, 990–993 (1999).
- Jones, K. A. *et al.* GABA_B receptors function as a heteromeric assembly of the subunits GABA_BR1 and GABA_BR2. *Nature* **396**, 674–679 (1998).
- White, J. H. *et al.* Heterodimerization is required for the formation of a functional GABA_B receptor. *Nature* **396**, 679–682 (1998).
- Kaupmann, K. *et al.* GABA_B-receptor subtypes assemble into functional heteromeric complexes. *Nature* **396**, 683–687 (1998).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Collaborative Computational Project, Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Weik, M. *et al.* Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci. USA* **97**, 623–628 (2000).
- Masu, M., Tanabe, Y., Tsuchida, K., Shigemoto, R. & Nakanishi, S. Sequence and expression of a metabotropic glutamate receptor. *Nature* **349**, 760–765 (1991).
- Flor, P. J. *et al.* Molecular cloning, functional expression and pharmacological characterization of the human metabotropic glutamate receptor type 2. *Eur. J. Neurosci.* **7**, 622–629 (1995).
- Minakami, R., Katsuki, F., Yamamoto, T., Nakamura, K. & Sugiyama, H. Molecular cloning and the functional expression of two isoforms of human metabotropic glutamate receptor subtype 5. *Biochem. Biophys. Res. Commun.* **199**, 1136–1143 (1994).
- Hashimoto, T. *et al.* The whole nucleotide sequence and chromosomal localization of the gene for human metabotropic glutamate receptor subtype 6. *Eur. J. Neurosci.* **9**, 1226–1235 (1997).
- Garrett, J. E. *et al.* Molecular cloning and functional expression of human parathyroid calcium receptor cDNAs. *J. Biol. Chem.* **270**, 12919–12925 (1995).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
- Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

Acknowledgements

We thank N. Yagi and Y. Katsuya for use of the facilities at SPring-8, Hyogo, Japan. We also thank T. Tomura and S. Yamamoto for technical assistance; H. Toh and T. Hiroike for their help in bioinformatic analyses; and Y. Shimura, H. Nakamura and A. Pähler for discussions and encouragement. This study is partly supported by a grant from the SPring-8 Joint Research Promotion Scheme of the Japan Science and Technology Corporation.

Correspondence and requests for materials should be addressed to K.M. (e-mail: morikawa@beri.co.jp) or H.J. (e-mail: jingami@beri.co.jp). Atomic coordinates have been deposited in the PDB under accession codes 1EWT for free form I, 1EWV for free form II, and 1EWK for the complex form.

Extremely red Kuiper-belt objects in near-circular orbits beyond 40 AU

S. C. Tegler* & W. Romanishin†

* Department of Physics & Astronomy, Northern Arizona University, Flagstaff, Arizona 86011, USA

† Department of Physics & Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA

Kuiper-belt objects (KBOs) are an ancient reservoir of comets beyond Neptune's orbit^{1,2}. Some of these objects were recently found to have the reddest optical colours in the Solar System³, but the number of objects for which accurate colours were available was too small for any correlation to be discerned between colour and physical or dynamical properties, which might shed light on the origin of these objects. Here we report that all nine of the KBOs in our survey on near-circular (low-eccentricity) orbits with perihelion distances larger than 40 AU have extremely red surfaces, thereby connecting an observable property with a dynamical class. Of the objects with orbital eccentricities greater than 0.1, about half are also very red, while the rest have colours similar to the Sun, meaning that reflected sunlight is not strongly modified by the objects' surface properties. In addition, of the 13 'classical' KBOs (those with semimajor axis $a \approx 45$ AU and eccentricity $e < 0.15$), the ten that are very red are in orbits with small angles of inclination to the ecliptic, whereas the three with solar colours are all in high-inclination orbits. We suggest that these three 'grey' classical KBOs may be part of a dynamical group that is separate from the 'red' classical KBOs.

Our survey uses the low-resolution imaging spectrometer (LRIS), B (438 nm), V (547 nm), and R (642 nm) glass filters, and a $2,048 \times 2,048$ CCD (charge coupled device) at the $f/15$ Cassegrain focus of the 10-m Keck II telescope⁴. Each of our images covers 6×7 arcmin on the sky at 0.21 arcsec per pixel and has a stellar point-spread function full-width at half-maximum of about 0.7 arcsec. We correct instrumental magnitudes to zero air mass using nightly extinction coefficients and place them on the Kron-Cousins system using transformation equations from observations of Landolt standard star fields⁵. A full description of our data reduction and analysis techniques can be found in the literature^{3,6-8}.

Table 1 presents our results. Measurements for 17 objects come from the Keck telescope, and measurements for four objects come

from the 1.8-m Vatican telescope and the 2.3-m University of Arizona telescope. In Fig. 1a, we plot 17 points from observations on the Keck telescope, two points from observations on the Vatican telescope, and 13 points from our original survey on the University of Arizona 2.3-m telescope³. Colours for KBOs and Centaurs (recent refugees from the Kuiper belt on outer planet-crossing orbits) are represented by filled and open squares, respectively. The colour of the Sun is represented by the standard solar symbol at $B - V = 0.67$ and $V - R = 0.36$. The reddest objects are toward the upper right-hand corner of Fig. 1a. The impact of the new data on our survey can be seen in the $B - R$ histogram in Fig. 1b. We obtain the $B - R$ colour for a KBO from the sum of the $B - V$ and $V - R$ colours. The Sun has a $B - R$ colour of 1.03. The reddest objects in Fig. 1b are in bins with the largest values of $B - R$. Values from our original survey and the current survey are represented by grey and white bars, respectively. Despite a spread in colour space by the current survey, we clearly see a double-peaked distribution in the $B - R$ histogram. Apparently, the surfaces of some KBOs and Centaurs are rich in a material that absorbs blue sunlight. The lack of blue light upon reflection gives the surfaces their red colour. Other KBOs and Centaurs have little or none of the blue absorber. The material on their surfaces absorbs blue, yellow and red sunlight with a similar efficiency and we refer to the surfaces of these objects as grey in colour.

Now that our survey has a significant number of KBOs and Centaurs, we can look for patterns between the colours of the objects and their orbital elements. In Fig. 2a, we plot $B - R$ colour against perihelion distance, q . The Centaurs (open squares) and KBOs (filled squares) are evenly split between red surfaces (top) and grey surfaces (bottom) for perihelion distances between 6 and 40 AU. Beyond 40 AU, all seven KBOs with $B - R$ colours reside in the red group. In addition to the seven KBOs with known $B - R$ colours, there are two more KBOs in Table 1 with perihelion distances beyond 40 AU that are in the red group only on the basis of their $V - R$ colours: these are 1997 CU29 and 1997 CT29. A statistical calculation suggests the probability of getting the first nine objects with perihelion distances greater than 40 AU all red in a population that is evenly split between red and grey colours (as it is for perihelion distances out to 40 AU) is 1 in 512. Evidently, KBOs with perihelion distances greater than 40 AU are dominated by extraordinarily red surfaces.

There is another possible pattern in Fig. 2a. Objects in the red group appear redder with smaller perihelion distance. Perhaps we are seeing the effect of an increase of solar radiation modifying the surfaces.

Table 1 Keck colours of Kuiper-belt objects and Centaurs

Object	Class	V	$\frac{\sigma}{\sqrt{n}}$	B - V	$\frac{\sigma}{\sqrt{n}}$	V - R	$\frac{\sigma}{\sqrt{n}}$	UT
1999 TR11	Plutino	23.51	0.07	1.02	0.08	0.75	0.07	1999 Oct 07
1999 OY3	Classical	22.53	0.02	0.71	0.01	0.37	0.02	1999 Sep 08
1998 WX24	Classical	23.44	0.04	1.09	0.05	0.70	0.05	1999 Oct 07
1998 WV24	Plutino	23.41	0.01	0.77	0.01	0.50	0.03	1999 Sep 08
1998 WH24*	Classical	21.23	0.01	0.94	0.03	0.62	0.03	1999 Nov 08-12
1998 SM165*	1:2	21.61	0.05	1.01	0.10	0.75	0.07	1999 Nov 13-15
1997 SZ10	1:2	23.69	0.03	1.14	0.08	0.65	0.03	1999 Oct 07
1997 QH4	Classical	23.70	0.04	1.01	0.07	0.67	0.05	1998 Oct 26
1997 CU29†	Classical	23.31	0.08			0.66	0.04	1998 Jan 23-26
1997 CT29†	Classical	23.72	0.11			0.75	0.12	1998 Jan 23-26
1996 TK66	Classical	23.14	0.03	0.99	0.02	0.63	0.02	1998 Oct 26
1996 SZ4	Plutino	23.33	0.03	0.83	0.03	0.52	0.02	1999 Sep 07
1996 RR20	Plutino	23.57	0.01	1.16	0.04	0.71	0.03	1999 Sep 07
1995 QZ9	Plutino	24.12	0.02	0.88	0.04	0.52	0.03	1999 Sep 07
1994 VK8	Classical	24.07	0.05	1.01	0.06	0.67	0.03	1998 Oct 26
1993 SB	Plutino	23.13	0.01	0.82	0.03	0.47	0.04	1999 Sep 08
1993 RO	Plutino	23.95	0.03	0.85	0.07	0.51	0.06	1999 Sep 08
1992 QB1	Classical	23.79	0.03	0.92	0.06	0.78	0.03	1999 Sep 08
1999 OX3	Centaur	22.45	0.03	1.12	0.03	0.69	0.01	1999 Sep 08
1998 QM107	Centaur	23.11	0.02	0.73	0.02	0.52	0.03	1999 Sep 07
1994 TA	Centaur	24.31	0.05	1.24	0.10	0.68	0.03	1998 Oct 26

* Vatican Observatory 1.8-m.

† University of Arizona 2.3-m.

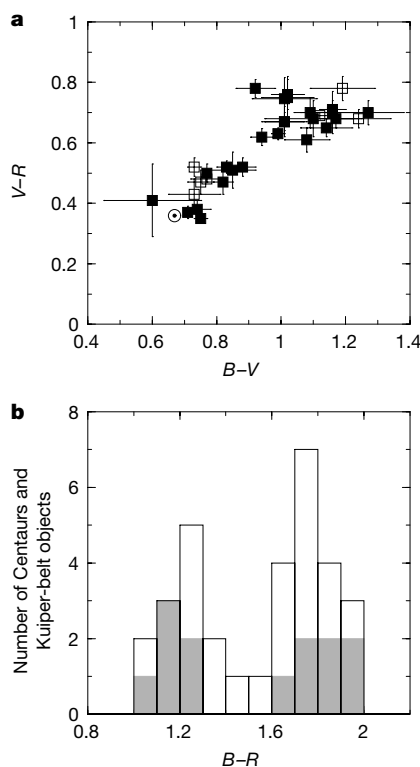


Figure 1 Surface colours of KBOs. **a**, A colour–colour plot of the KBOs (solid squares) and Centaurs (open squares) in our survey. The colour of the Sun is represented by the standard solar symbol at $B - V = 0.67$ and $V - R = 0.36$. The error bars are the uncertainty in the mean, $\sigma/\sqrt{n}$, where σ is the standard deviation of n 300-second exposures. Photon noise, sky noise, and ‘unseen’ background objects just below our detection limit are sources of random error that contribute to the size of the error bars. Thirteen points come from our initial survey³, seventeen points come from recent observations on the 10-m Keck telescope, and two points come from recent observations on the 1.8-m Vatican telescope. The reddest objects are toward the upper right-hand corner. **b**, A $B - R$ histogram of the 32 objects in our survey that comes from the sum of their $B - V$ and $V - R$ colours. The grey and white bars represent our initial and current surveys, respectively. The reddest objects are in bins with the largest values of $B - R$. There is a clear double-peaked distribution of colours.

Besides patterns in perihelion distance, we can see some important trends in a plot of eccentricity, e , against semimajor axis, a . In Fig. 2b, we see that our survey contains 13 KBOs, Plutinos, on stable Pluto-like orbits in a 2:3 resonance with Neptune ($a \approx 39.5$ AU and $0.1 < e < 0.3$) and 13 classical KBOs on distant and near-circular orbits ($a \approx 45$ AU and $e < 0.15$). Objects with perihelion distances at and beyond 40 AU fall on and below the curve toward the lower right-hand corner of Fig. 2b. In our survey we find four red and four grey Centaurs, six red and seven grey Plutinos, and ten red and three grey classical KBOs. Unlike the Centaurs and Plutinos which are evenly split between red and grey colours, the classical KBOs are dominated by red colours. The dominance of red surfaces among the classical KBOs is due to the nine KBOs with perihelion distances greater than 40 AU.

The ten classical KBOs with red colours share a trait that the three classical KBOs with grey colours do not share. In particular, the ten red objects are all on orbits with small inclination angles to the ecliptic, less than 13° . On the other hand, the three grey objects are all on orbits with large inclination angles, greater than 24° . The probability of getting ten classical KBOs with the lowest inclination orbits all red in a population that is evenly split between red and grey colours is 1 in 1,024. Perhaps the three high-inclination grey KBOs currently classified as classical KBOs are really members of a separate dynamical group.

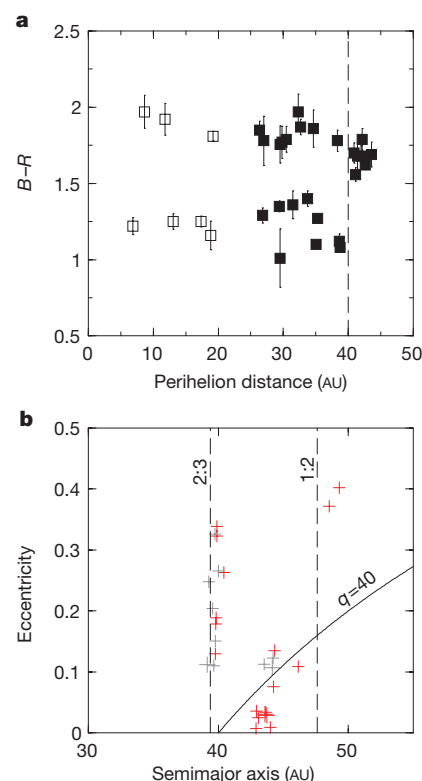


Figure 2 Objects on circular orbits beyond 40 AU are red. **a**, A plot of $B - R$ colour against perihelion distance for the objects in our survey. The uncertainties in q are smaller than the size of the symbols. The objects with perihelion distances between 6 and 40 AU are evenly split between the red (larger $B - R$ values) and grey (smaller $B - R$ values) groups, but all seven KBOs with perihelion distances beyond 40 AU and $B - R$ colours are red. **b**, A plot of orbital eccentricity against semimajor axis. Red and grey symbols represent objects with red and grey surface colours, respectively. Although the Plutinos ($a \approx 39.5$ AU) are evenly split between the red and grey groups, the classical KBOs ($a \approx 45$ AU) are dominated by red surface colours. All nine classical KBOs in our survey with perihelion distances greater than 40 AU, below the curve toward the lower right-hand side of the figure, are red.

Another noteworthy point concerns the two red KBOs in the 1:2 resonance with Neptune in Fig. 2b ($a \approx 48$ AU and $e \approx 0.4$). KBOs that are currently in the 2:3 and 1:2 resonances with Neptune may initially have been on circular orbits closer to the Sun, but as Neptune migrated outward, it swept them outward onto eccentric orbits⁹. Perhaps the two objects in the 1:2 resonance in Fig. 2b were initially on near-circular orbits among the classical KBOs with red surface colours. □

Received 24 July; accepted 11 September 2000.

- Jewitt, D. & Luu, J. Discovery of the candidate Kuiper belt object 1992 QB1. *Nature* **362**, 730–732 (1993).
- Marsden, B. G. List of transneptunian objects. (Cited June 2000) <<http://cfa-www.harvard.edu/cfa/ps/lists/TNOs.html>> (2000).
- Tegler, S. C. & Romanishin, W. Two distinct populations of Kuiper-belt objects. *Nature* **392**, 49–51 (1998).
- Oke, J. B. *et al.* The Keck Low Resolution Imaging Spectrometer. *Publ. Astron. Soc. Pacif.* **107**, 375–385 (1995).
- Landolt, A. U. UBVR photometric standard stars in the magnitude range $11.5 < V < 16.0$ around the celestial equator. *Astron. J.* **104**, 340–371 (1992).
- Romanishin, W., Tegler, S. C., Levine, J. & Butler, N. BVR photometry of centaur objects 1995 GO, 1993 HA2, and 5145 Pholus. *Astron. J.* **113**, 1893–1898 (1997).
- Tegler, S. C. & Romanishin, W. The extraordinary colors of trans-neptunian objects 1994 TB and 1993 SC. *Icarus* **126**, 212–217 (1997).
- Tegler, S. C., Romanishin, W., Stone, A., Tryka, K., Fink, U. & Fevig, R. Photometry of the trans-neptunian object 1993 SC. *Astron. J.* **114**, 1230–1233 (1997).
- Malhotra, R. The origin of Pluto's orbit: implications for the solar system beyond Neptune. *Astron. J.* **110**, 420–429 (1995).

Acknowledgements

We thank the NASA Planetary Astronomy programme for financial support of this research and the NASA Keck Telescope Allocation Committee for consistent allocation of telescope time.

Correspondence and requests for materials should be addressed to S.C.T. (e-mail: tegler@proto.phy.nau.edu).

n-type colloidal semiconductor nanocrystals

Moonsub Shim & Philippe Guyot-Sionnest

James Franck Institute, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637, USA

Colloidal semiconductor nanocrystals^{1,2} combine the physical and chemical properties of molecules with the optoelectronic properties of semiconductors. Their colour is highly controllable, a direct consequence of quantum confinement on the electronic states³. Such nanocrystals are a form of 'artificial atoms' (ref. 4) that may find applications in optoelectronic systems such as light-emitting diodes^{5,6} and photovoltaic cells⁷, or as components of future nanoelectronic devices. The ability to control the electron occupation (especially in n-type or p-type nanocrystals) is important for tailoring the electrical and optical properties, and should lead to a wider range of practical devices. But conventional doping by introducing impurity atoms has been unsuccessful so far: impurities tend to be expelled from the small crystalline cores (as observed for magnetic impurities⁸), and thermal ionization of the impurities (which provides free carriers) is hindered by strong confinement. Here we report the fabrication of n-type nanocrystals using an electron transfer approach commonly employed in the field of conducting organic polymers⁹. We find that semiconductor nanocrystals prepared as colloids can be made n-type, with electrons in quantum confined states.

Electron transfer in and out of nanocrystals has been a subject of study for many years, mostly in the context of photochemistry and photovoltaics¹⁰. However, there has never been any conclusive evidence that injected electrons could be placed in the Lowest Unoccupied Quantum-Confined Orbital (we will use the acronym LUQCO) which is the essence of making n-type nanocrystals. Instead, the electrons were most probably transferred to trap states. Reducing the nanocrystals such that they are n-type requires the absence or saturation of electron traps and a slow kinetics towards oxidation. This may seem hopeless given the large surface area of the nanocrystals and the possible presence of many dangling bonds. However, the conduction band minimum of many semiconductors lies near the reduction potential of hydrogen¹¹, well below the reduction potential of alkali metals (for example, the reduction potential of Na = -2.71 V versus SHE). While quantum confinement shifts the energy levels of the conduction band higher (relative to the bulk), that shift is typically less than 1 eV for most sizes of quantum dots studied. Therefore it should be feasible to inject electrons into semiconductor nanocrystals with alkali metals in a manner similar to organic polymers⁹, C₆₀ (ref. 12) and carbon nanotubes¹³. Using surface-passivated nanocrystals in anhydrous and de-aerated solutions, we have observed successful electron transfer from sodium biphenyl to the LUQCO of colloidal semiconductor nanocrystals, making them the first n-type nanocrystals. The radical anion of biphenyl acts as the charge shuttle. The n-type nanocrystals can also be obtained by simply exposing a solution of

nanocrystals to chunks of sodium. However, this process typically takes several days owing to the low mobility of the nanocrystals.

To ascertain that the electrons are successfully placed in the LUQCO of the nanocrystals, rather than merely charging them by occupation of surface states, infrared spectroscopy of the 1S_c–1P_c intraband transition has been used. As seen in the inset of Fig. 1, if an electron is successfully transferred into the LUQCO (labelled 1S_c for a spherical nanocrystal), a strong electronically allowed infrared transition to the next higher state (labelled 1P_c) must appear. At the same time, a bleach of the exciton transition must occur as a result of the occupation of the 1S_c state. Therefore a combination of infrared and optical spectroscopy provides a rigorous diagnosis of the n-type character of the nanocrystals.

In Fig. 1, infrared and visible absorption spectra of 5.2-nm CdSe nanocrystals before (dotted line) and about 1 minute after (solid line) the addition of sodium biphenyl reagent are shown. Upon charge transfer, the first and the second exciton peaks at 2.07 and 2.18 eV, respectively, are strongly bleached and broadened. Broadening is expected if there are charges that can shift the exciton energy by the Stark effect and this can also lead to an apparent blue shift. As the charges may reside in surface states and/or in the delocalized 1S_c state, the changes in the visible absorption indicate that electron transfer may possibly have occurred but do not guarantee that the nanocrystals are n-type. The true n-type character of the nanocrystals is unambiguously confirmed by the appearance of the 1S_c–1P_c infrared absorption arising at 0.3 eV.

The size-dependence of this 1S_c–1P_c transition in the n-type nanocrystals is similar to photoexcited nanocrystals^{14,15}. Figure 2 shows infrared absorption spectra corresponding to the 1S_c–1P_c transition in n-type CdSe nanocrystals of different sizes capped with trioctylphosphine oxide (TOPO). Figure 3 compares the size dependence of the 1S_c–1P_c transition observed in n-type CdSe nanocrystals (filled circles) to the effective mass calculation¹⁶ (solid line) and shows a fair agreement. A previously reported 1S_c–1P_c transition observed in optically excited CdSe nanocrystals¹⁴ (open triangles) as well as for n-type CdSe nanocrystals with various caps (filled diamonds) are also shown in Fig. 3. As we expected, whether

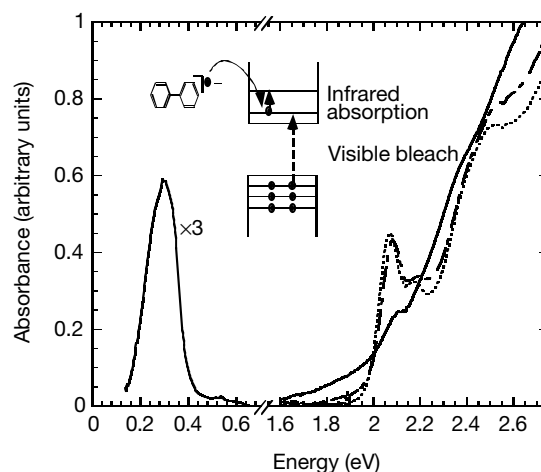


Figure 1 Absorption spectra of CdSe nanocrystals. Spectra are shown before (dotted line), immediately after (solid line), and 27 hours after (dashed line) the addition of sodium biphenyl reagent. The concurrent optical bleach of the first two exciton transitions and the appearance of the infrared absorption are clearly seen. The blue-shift of the optical spectra after the disappearance of the infrared absorption suggests that the n-type nanocrystals decompose by loss of the outermost layer of the semiconductor. Typically, n-type nanocrystals are made by addition of sodium biphenyl to a dried and degassed solution of nanocrystals in heptamethylnonane (~0.1 mM in nanocrystals and ~50 mM in sodium biphenyl). Small amounts of TOPO (<5 mg ml⁻¹) can be added to prevent precipitation.

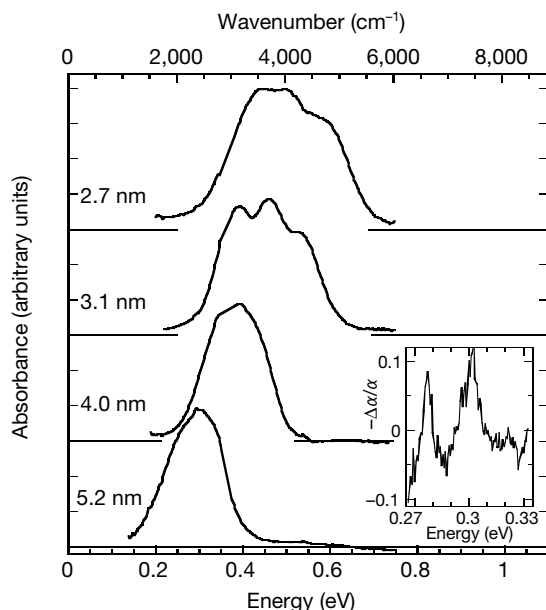


Figure 2 Infrared absorption spectra of n-type CdSe nanocrystals capped with TOPO. The sizes correspond to the diameter of the core semiconductor. Between 0.35 to 0.37 eV, the solvent C-H stretch blocks out the infrared source and the absorbance in this range is omitted for clarity. The $1S_e-1P_e$ transitions in the smallest two sizes of n-type CdSe nanocrystals show three distinct features, possibly arising from the lifting of threefold degeneracy of $1P_e$ states. The inset is the infrared hole-burning spectrum of 5.2-nm samples at 20 K. The main bleach is at the pump frequency, 0.3 eV, and the strong phonon replica appears at 0.275 eV. The longitudinal optical (LO) phonon for bulk CdSe is 26 meV. α , absorbance.

the surface capping group introduces hole traps (for example, 4-methoxythiophenol) or removes these traps (for example, ZnS coating), the electron injection process to produce n-type semiconductor nanocrystals is not affected.

The $1S_e-1P_e$ transitions are inhomogeneously broadened mostly from a distribution of sizes ($\sigma \approx 5$ to 8%). The linewidths are about 30% larger than that of the first exciton peak in the visible spectra of the neutral nanocrystals (for samples shown in Fig. 2, the infrared full-widths at half-maximum (FWHM) are 0.26, 0.22, 0.15 and 0.14 eV and the visible FWHM values are 0.19, 0.16, 0.14 and 0.11 eV from the smallest to the largest sizes, respectively). Low-temperature infrared hole-burning spectra of n-type CdSe nanocrystals confirm that the linewidths are dominated by the size distribution. At 20 K, an upper limit of 5 meV is obtained for the homogeneous linewidth (M.S. and P.G.-S., unpublished work) of 5.2-nm nanocrystals (see Fig. 2 inset).

The $1S_e-1P_e$ transition in n-type CdSe nanocrystals presents a very large absorbance, comparable to the first exciton transition. In Fig. 1, it is nearly one half that of the first exciton transition in the visible. Infrared absorbance almost as large as the first exciton transition has been observed in 5.4-nm CdSe nanocrystals (absorbance at 2.05 eV was 0.423 before sodium biphenyl addition compared to peak absorbance at 0.3 eV, which was 0.395). Using the molar extinction coefficients of the visible absorption spectra¹⁷, the infrared molar absorptivity for this sample is about $0.8 \times 10^6 \text{ M}^{-1} \text{ cm}^{-1}$.

The long-term stability of the n-type nanocrystals is strongly affected by the presence of impurities, exposure to air and the dryness of the solution. The decay is probably due to further oxidation and decomposition. In n-type CdSe nanocrystals, the $1S_e-1P_e$ transition decays within 30 minutes to 24 hours (from the smallest to the largest sizes, respectively) at room temperature. If the solution of nanocrystals and biphenyl anions is intentionally

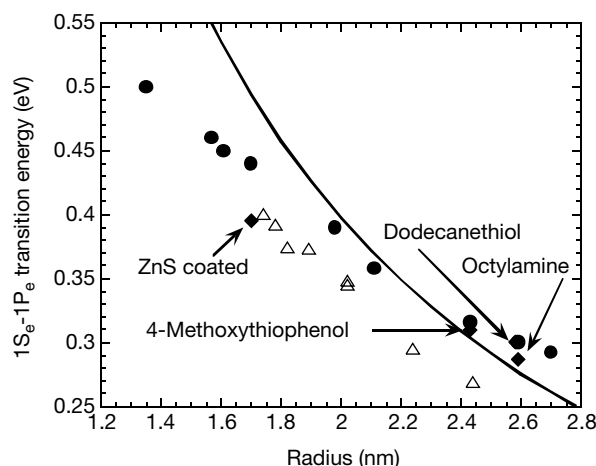


Figure 3 Experimental $1S_e-1P_e$ transition energy of n-type CdSe nanocrystals. The transition energies of the nanocrystals capped with TOPO (filled circles) and optically excited nanocrystals (open triangles) are compared to effective mass calculation (solid line). The filled diamonds are for n-type CdSe nanocrystals with various capping groups as indicated. The infrared absorption band is easily tuned with size and its position can be readily determined from bulk properties.

exposed to air, decay occurs within minutes. The stability of the n-type nanocrystals is also strongly temperature-dependent because at 20 K the samples appear to be stable indefinitely (in excess of five weeks). Several hours to days after the complete decay of the infrared absorption, there is a recovery of the optical bleach. After recovery, the nanocrystals are of smaller sizes as indicated by the blue-shift of the optical absorption features shown in Fig. 1 (dashed line). The blue-shift from 2.07 to 2.08 eV corresponds to a reduction in size of about 2 Å in diameter. We note that the infrared absorption and the optical bleach can be 'turned on' again by the addition of more sodium biphenyl reagent.

If no excess capping molecules (TOPO and trioctylphosphine (TOP) for samples in Figs 1 and 2) are present in the solution, a slow precipitation of the nanocrystals occurs. We note that excess TOPO (without excess TOP) present in solution prevents this slow precipitation. As the reduction potential of Cd/Cd^{2+} is -0.403 V versus the standard hydrogen electrode (SHE) near or below the LUQCO of CdSe nanocrystals, the most likely pathway for the decomposition of n-type CdSe nanocrystals is by oxidation with loss of electrons in Cd/TOP complexes in solution. As expected, smaller n-type nanocrystals with a larger degree of confinement are less stable. Determining which systems can be made n-type by the electron transfer method and the thermodynamic stability of the n-type nanocrystals may simply be approached by looking at the reduction potential of the nanocrystals and of their constituent elements. For example, n-type ZnSe nanocrystals may be less stable because the reduction potential of Zn/Zn^{2+} is -0.762 V versus SHE whereas the conduction band minimum of bulk ZnSe is near -1.5 V versus SHE¹¹. On the other hand, ZnO, with its valence band minimum near the reduction potential of hydrogen¹¹, should be an excellent candidate. In fact, charge transfer to ZnO nanocrystals has been previously investigated by electrochemistry but the ultraviolet-visible spectra could not distinguish between electron transfer to trap states or to the LUQCO^{18,19}.

Figure 4 shows the infrared absorption spectra of n-type ZnO, CdS and CdSe nanocrystals prepared by the electron transfer method. The visible bleach and tail are also observed in all three types of nanocrystals. The relatively broader infrared absorption band of ZnO may be attributed to a larger size distribution ($\sigma \approx 25\%$) as well as to a stronger electron-phonon interaction. The insets of Fig. 4 are the corresponding time evolution of the

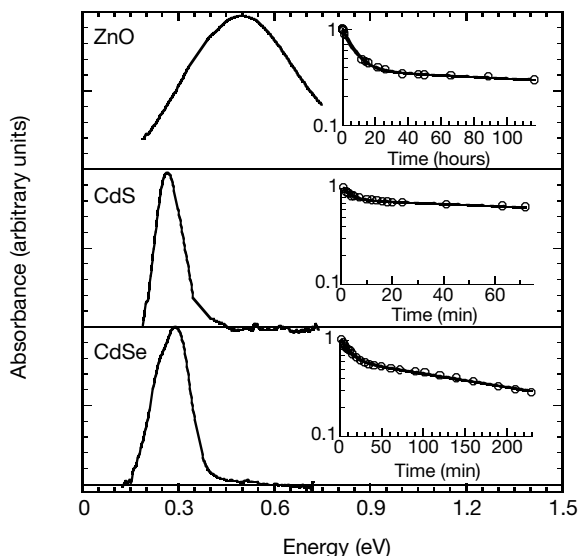


Figure 4 Infrared absorption spectra of n-type ZnO (4.4 nm), CdS (7 nm) and CdSe (5.4 nm) nanocrystals capped with TOPO. The insets are the corresponding time evolution of the infrared absorbance maximum normalized for comparison. We note that the x-axis of the inset for ZnO is in hours.

infrared absorption maximum. As expected from the relative reduction potentials of the semiconductor and of the constituent elements, n-type ZnO nanocrystals are more stable than CdS and CdSe. At room temperature, approximately 30% of the initial infrared absorption is observed 5 days after the initial preparation of n-type ZnO nanocrystals, whereas the infrared absorption completely decays in less than 2 days for both CdS and CdSe nanocrystals. The intrinsic limit of this electron injection method should be the point at which the sum of the confinement energy, the charging energy and the position of the conduction band minimum reaches the reduction potential of the reducing species.

One of the advantageous properties of quantum dots prepared as colloids is the organic capping layer. The readily exchangeable organic surfactants allow for a versatile manipulation of nanocrystals in many different environments. For example, recapping nanocrystals with organic functional groups that are compatible with physiological environments has shown that nanocrystals can be useful biological tags^{20,21}. The electron injection into the LUQCO is also achieved in nanocrystals with different capping groups as indicated in Fig. 3. As long as the capping layer does not introduce electron traps within the band gap, this method should be applicable. The capping layer also provides an avenue to improve the long-term stability of n-type nanocrystal (for example, binding of the cation to the surface and core/shell structures).

Colloidal semiconductor nanocrystals of various materials (CdSe, CdS and ZnO) can be reduced to n-type by Na or biphenyl radical anions with electrons occupying the quantum confined states of the conduction band. A stability of hours to days at room temperature is observed and improvements are likely in the future. We have shown how electron occupation of the $1S_e$ state dramatically affects the optical properties, creating the possibility for strong electrochromic response in the visible and mid-infrared. The conductivity of films of n-type nanocrystals is one of the most interesting aspects to investigate, owing to the possibility of enhanced inter-nanocrystal electron transfer that may lead to photovoltaic or electronic applications. In general, the control of the Fermi level should be important in future applications of colloidal semiconductor nanocrystals and the electron transfer method may be the most viable approach in the nanometer length scale with strong confinement. □

Received 22 June; accepted 6 September 2000.

1. Nirmal, M. & Brus, L. E. Luminescence photophysics in semiconductor nanocrystals. *Acc. Chem. Res.* **32**, 407–414 (1999).
2. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
3. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
4. Ashoori, R. C. Electrons in artificial atoms. *Nature* **379**, 413–419 (1996).
5. Colvin, V. L., Schlamp, M. C. & Alivisatos, A. P. Light-emitting diodes made from cadmium selenide nanocrystals and a semiconducting polymer. *Nature* **370**, 354–357 (1994).
6. Dabbousi, B. O., Bawendi, M. G., Onitsuka, O. & Rubner, M. F. Electroluminescence from CdSe quantum-dot/polymer composites. *Appl. Phys. Lett.* **66**, 1316–1318 (1995).
7. O'Regan, B. & Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. *Nature* **353**, 737–740 (1991).
8. Mikulec, E. V. *et al.* Organometallic synthesis and spectroscopic characterization of manganese-doped CdSe nanocrystals. *J. Am. Chem. Soc.* **122**, 2532–2540 (2000).
9. Chiang, C. K. *et al.* Synthesis of highly conducting films of derivatives of polyacetylene, CH_x . *J. Am. Chem. Soc.* **100**, 1013–1015 (1978).
10. Kamat, P. V. Interfacial charge transfer processes in colloidal semiconductor systems. *Prog. Reaction Kinetics* **19**, 277–316 (1994).
11. Miller, R. J. D., McLendon, G. L., Nozik, A. J., Schmickler, W. & Willig, F. *Surface Electron Transfer Processes* (VCH Publishers, New York, 1995).
12. Haddon, R. C. *et al.* Conducting films of C_{60} and C_{70} by alkali-metal doping. *Nature* **350**, 320–322 (1991).
13. Lee, R. S., Kim, H. J., Fischer, J. E., Thess, A. & Smalley, R. E. Conductivity enhancement in single-walled carbon nanotube bundles doped with K and Br. *Nature* **388**, 255–257 (1997).
14. Guyot-Sionnest, P. & Hines, M. A. Intraband transitions in semiconductor nanocrystals. *Appl. Phys. Lett.* **72**, 686–688 (1998).
15. Shim, M., Shilov, S. V., Braiman, M. S. & Guyot-Sionnest, P. Long-lived delocalized electron states in quantum dots: a step-scan Fourier transform infrared study. *J. Phys. Chem.* **104**, 1494–1496 (2000).
16. Norris, D. J. & Bawendi, M. G. Measurement and assignment of the size-dependent optical spectrum in CdSe quantum dots. *Phys. Rev. B* **53**, 16338–16346 (1996).
17. Shim, M. & Guyot-Sionnest, P. Permanent dipole moment and charges in colloidal semiconductor quantum dots. *J. Chem. Phys.* **111**, 6955–6964 (1999).
18. Hoyer, P., Eichberger, R. & Weller, H. Spectroelectrochemical investigations of nanocrystalline ZnO films. *Ber. Bunsenges. Phys. Chem.* **97**, 630–635 (1993).
19. Hoyer, P. & Weller, H. Size-dependent redox potentials of quantized zinc oxide measured with an optically transparent thin layer electrode. *Chem. Phys. Lett.* **221**, 379–384 (1994).
20. Bruchez, M. Jr, Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
21. Chan, W. C. W. & Nie, S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**, 2016–2018 (1998).

Acknowledgements

This work was funded by the US NSF. We made use of the Materials Research Science and Engineering Center (MRSEC) shared facilities supported by the US NSF.

Correspondence and requests for materials should be addressed to M.S. (e-mail: mshim@uchicago.edu) or P.G.S. (e-mail: pgs@uchicago.edu).

Three-dimensional control of light in a two-dimensional photonic crystal slab

Edmond Chow*, S.Y. Lin*, S.G. Johnson†, P.R. Villeneuve†, J.D. Joannopoulos†, J.R. Wendt*, G.A. Vawter*, W. Zubrzycki*, H. Hou* & A. Alleman*

* Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA

† Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Optoelectronic devices are increasingly important in communication and information technology. To achieve the necessary manipulation of light (which carries information in optoelectronic devices), considerable efforts are directed at the development of photonic crystals—periodic dielectric materials that have so-called photonic bandgaps, which prohibit the propagation of photons having energies within the bandgap region. Straightforward application of the bandgap concept is generally thought to

require three-dimensional (3D) photonic crystals^{1–5}; their two-dimensional (2D) counterparts confine light in the crystal plane^{6,7}, but not in the perpendicular z direction, which inevitably leads to diffraction losses. Nonetheless, 2D photonic crystals still attract interest^{8–15} because they are potentially more amenable to fabrication by existing techniques and diffraction losses need not seriously impair utility. Here we report the fabrication of a waveguide-coupled photonic crystal slab (essentially a free-standing 2D photonic crystal) with a strong 2D bandgap at wavelengths of about $1.5\ \mu\text{m}$, yet which is capable of fully controlling light in all three dimensions. These features confirm theoretical calculations^{16,17} on the possibility of achieving 3D light control using 2D bandgaps, with index guiding providing control in the third dimension, and raise the prospect of being able to realize unusual photonic-crystal devices, such as thresholdless lasers¹.

The photonic crystal slab consists of cylindrical holes etched through a thin GaAs slab and partially into a $1\text{-}\mu\text{m}$ -thick underlying $\text{Al}_{0.9}\text{Ga}_{0.1}\text{As}$ layer (Fig. 1). The nearest hole-to-hole spacing is a and the hole diameter is d ($=0.6a$). The thickness (t) of the GaAs slab was chosen to be $0.5a$ to create a large photonic bandgap¹⁶. The $\text{Al}_{0.9}\text{Ga}_{0.1}\text{As}$ layer was then wet oxidized to give a layer of Al_xO_y . The latter has a low refractive index ($n \approx 1.5$; ref. 18), which helps to confine light in the GaAs slab.

In a periodic dielectric structure, electromagnetic waves experience a periodic scattering potential, as described by Maxwell's equations. Light is coherently scattered by this potential, much as electrons are scattered in an electronic crystal (Bragg scattering). So the language for describing electronic states in an electronic

crystal¹⁹—such as ‘crystal wavevector’ (K) and ‘Brillouin zone’—may also be applied to describe photonic states in a photonic crystal. As a result of Bragg reflection, frequency gaps can also form; within these gaps, the photonic density of states vanishes and the propagation of light is strictly forbidden in all directions within the crystal.

Figure 2 shows such a frequency dispersion (that is, ω versus k) of a photonic crystal slab. Here, ω and k are the frequency and the crystal wavevector associated with an electromagnetic wave, respectively. The shaded area represents the light-cone region—where leaky modes occur—and the unshaded region shows where guided modes occur: the connected dots represent the guided modes within the slab (Fig. 2a). The light-cone boundary is identified only in full 3D band-structure calculations¹⁷; its slope is determined by the refractive index of Al_xO_y , cladding layer, and is slightly nonlinear because of the small penetration of etched holes into this layer. Although the dielectric–air interface of an individual hole would give rise to strong scattering and radiation losses, a periodic hole array causes coherent scattering, so that energy-momentum conservation rules prohibit the Bloch guided modes from coupling into air modes.

Guided modes may be even (open circles in Fig. 2a) or odd (solid circles) with respect to reflection through the mirror plane (x – y plane in Fig. 1b) bisecting the slab and the cylinders of the 2D hole¹⁷. These even and odd states have strong similarities with respectively transverse electric (TE) and transverse magnetic (TM) states in two dimensions. For TE modes, there exists a large fundamental TE photonic bandgap at $\omega = 0.263$ – 0.34 . There is also a higher-order TM gap at $\omega = 0.345$ – 0.378 . Operating within these bandgaps, light can be strongly reflected by the 2D gaps and also index-guided vertically. This is how light can be fully controlled using a 2D photonic-crystal slab.

In Fig. 2b, the dispersion for an unoxidized photonic-crystal slab is shown. Here the slope of the light cone, c/n , is set by the refractive index of the $\text{Al}_{0.9}\text{Ga}_{0.1}\text{As}$ layer, $n = 2.9$; a higher n implies a lower light-cone boundary. Consequently, only the lowest TE and TM band are guided, and only a small photonic bandgap is expected. The use of a low-index Al_xO_y layer in our 2D slab structure design is essential for obtaining 2D guided modes and bandgaps.

To probe the intrinsic optical properties of a photonic-crystal slab, we performed transmission measurements along Γ – K . Diode lasers were used as the light source. The transmitted light was split, and fed into a calibrated InGaAs photodetector for intensity measurements, and to an infrared (IR) camera for imaging. For

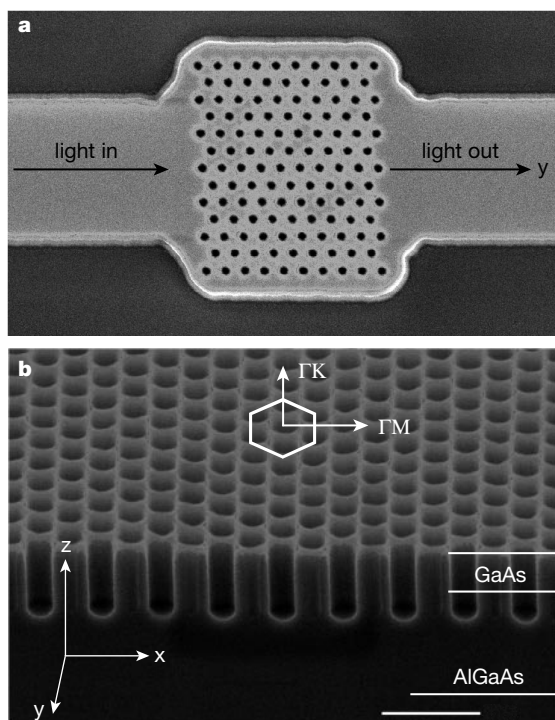


Figure 1 SEM images of a nanofabricated 2D photonic crystal slab. **a**, Top view showing an input ridge waveguide, a nine-period 2D hole array and an output waveguide. The holes are arranged in a triangular array, and the waveguides are used to facilitate light coupling. The 2D hole array section is a few rows wider than the waveguide to reduce light leakage around the side edges of photonic-crystal slab. **b**, A side view of the etched cylindrical holes, which have an etched depth of $\sim 0.5\ \mu\text{m}$ and side walls straight to within 5° . Nanometre-scale fabrication of 2D holes is done using a combination of electron-beam lithography and reactive-ion-beam etching processes. The major crystal symmetry directions, Γ – K and Γ – M , are also shown. Scale bar, $1\ \mu\text{m}$.

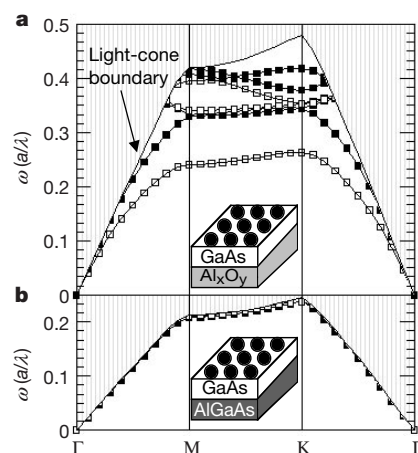


Figure 2 Computed dispersion of 2D photonic crystal slab structures. Here, ω is expressed in units of (a/λ) , and k is plotted along symmetry directions, Γ , $M(\pi/a)$ and $K(1.15\pi/a)$. **a**, The bottom cladding layer is Al_xO_y , with refractive index $n \approx 1.5$. See text for discussion of the light-cone boundary and the shaded area. **b**, The bottom cladding layer is $\text{Al}_{0.9}\text{Ga}_{0.1}\text{As}$ ($n = 2.9$).

precise optical alignment, the sample and lenses were mounted on a five-axis moving stage, which has a movement precision of better than 50 nm.

To obtain reliable transmittance data, the modal profile of the transmitted light must be carefully examined. In our experiment, Fig. 1b, the transmitted signal results from light coupling from (1) the input waveguide mode into (2) the photonic-crystal state and then back to (3) the output waveguide mode. If laser light is focused into the input waveguide and the crystal slab does not strongly scatter light, the output signal is a well-defined gaussian-shaped waveguide mode. But if the focus is inaccurate and the laser light is coupled, instead, into the undesired air mode or the substrate leaky mode, the output signal is typically broad and scattered in shape. The imaging camera at the output end is used to check the modal profile. Figure 3a (top) shows an image of TM light, $\lambda = 1,550$ nm, transmitted through a nine-period 2D crystal sample ($a = 460$ nm). The mode is bright and well defined, and its intensity profile (dashed line, bottom) is gaussian. The same measurement was repeated with TE light (Fig. 3b). The image is not as bright, suggesting the existence of a TE gap, but its profile remains gaussian. The observed gaussian profiles suggest that the output signal is a true measure of waveguide–crystal interaction.

To find the absolute intrinsic transmittance of a 2D crystal, a reference transmission is taken from an identical waveguide with no 2D crystal built in the middle section. By ratioing transmission signals, taken with and without a 2D crystal, absolute intrinsic transmittance is obtained. This procedure eliminates external uncertainties associated with reflection at waveguide–crystal interfaces and free-space-to-waveguide coupling efficiency.

In Fig. 4 we show (filled squares) the TM absolute transmittance along Γ –K versus the number of periods (N) of the hole array ($a = 460$ nm). The transmittance is high, (88 ± 7 %), and independent of N . The high transmittance shows that TM light is well-guided within the crystal slab, as predicted by the calculation shown in Fig. 2b. The observed N -independence suggests that radiation loss is not important for our samples, as such loss would give an exponential dependence on N . The small deviation from perfect 100% transmission may be due to scattering at the waveguide–photonic crystal interface, which is independent of N . This observation confirms the prediction that, under the light cone, light is guided in the 2D slab and vertical leakage loss is negligible.

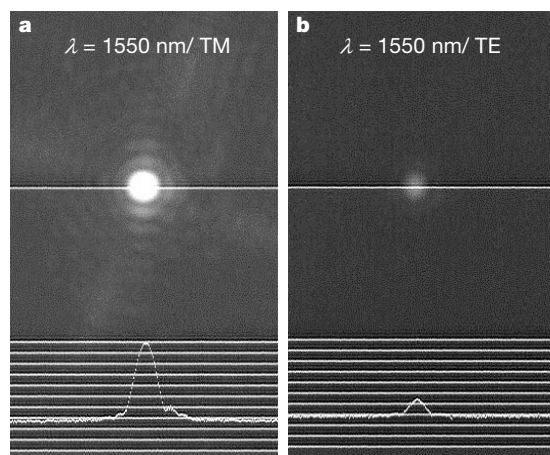


Figure 3 Infrared images of the transmitted laser light, and their corresponding modal profiles at $\lambda = 1,550$ nm. The input laser light is coupled into and out of a ridge waveguide using a pair of aspheric lenses with a high numerical aperture ($NA = 0.4$). An infrared camera is used to image the laser output. **a**, The image of TM transmitted light is brighter as it belongs to a guided mode. **b**, The TE transmitted light is much weaker and is in the photonic bandgap spectral regime. Both TM and TE modal profiles are gaussian-like. The horizontal lines are guides to the eye.

For the TE input at $\omega = 0.297$, light is strongly attenuated, and also slowly converted to TM polarization. The total and TE transmitted light is plotted in Fig. 4 as open and solid circles, respectively. The TE transmittance drops exponentially from 50% at $N = 1$, to about 0.2% at $N = 5$, and eventually saturates at about 6×10^{-4} . The observed exponential dependence shows that, at $\omega = 0.297$, the TE mode is in the photonic bandgap regime, which agrees with the prediction of Fig. 2a. As this TE mode is well below the light-cone boundary, vertical light leakage within the crystal is negligible. Thus, the observed absolute transmittance is a true measure of light attenuation in the 2D crystal slab. Light intensity is reduced ten times for every two periods it traverses in the crystal slab (or about 5 dB per period).

In Fig. 5a, the measured and computed TE transmission spectrum along Γ –K is plotted on a semi-log scale as open dots and a solid line, respectively. The theoretical transmittance was calculated using 3D finite difference time-domain simulations²⁰. The slightly larger observed bandgap, $\sim 8\%$, may be due to small uncertainties in the fabricated hole size and lattice constant. In the bandgap, $\omega \approx 0.27$, transmittance as low as about 2×10^{-4} is observed. At the lower band edge, $\omega \approx 0.25$, transmittance increases from 2×10^{-4} to unity over a small $\Delta\omega \approx 0.02$, a rise of four orders of magnitude. The upper and lower TE band edges occur at $\omega_1 \approx 0.34$ and $\omega_2 \approx 0.25$, respectively, yielding a large gap-to-midgap ratio, 30%. The close agreement between theory and experiment also confirms that light attenuation in the gap is due to photonic bandgap rejection, rather than to vertical leakage.

In Fig. 5b, the measured TM transmission spectrum is plotted as solid dots on a semi-log scale. A theoretical curve (solid line) is also shown, which again agrees well with the experimental data. In the allowed band (ω ranging from 0.240 to 0.300 a/λ), an overall high transmission of 70–100% is observed; this indicates that light is indeed well guided within the 2D slab for both TE and TM polarizations in the allowed bands.

The polarization conversion mentioned in Fig. 4 may be attributed to symmetry breaking in the 2D slab structure. In our sample, the upper cladding is essentially air, whereas the lower cladding is a 1- μm -thick Al_2O_3 layer. This asymmetry¹⁷ introduces a weak coupling between the otherwise non-interacting TE and TM modes. At $\omega = 0.297$, while TE light is attenuated due to the TE gap, TM light is free to propagate and its intensity should remain roughly constant. Thus for small N (< 3) in Fig. 4, TE light

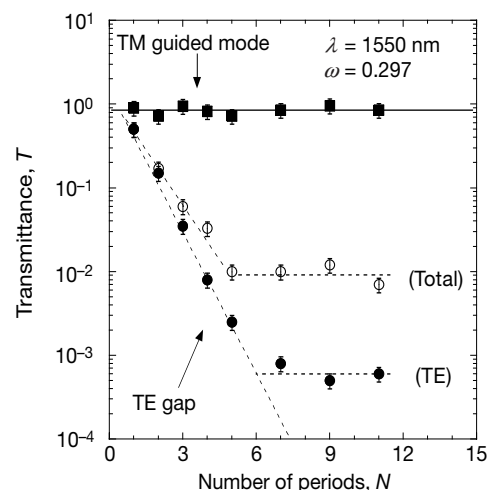


Figure 4 Absolute transmittance (T) at $\omega = 0.297$ taken from a series of samples with different number of periods (N). The transmittance for TM input is shown as filled squares. For TE input, transmitted light contains both TE and TM components. Both the TE (solid circles) and total (open circles) transmitted light attenuates exponentially as a function of N . See text for details.

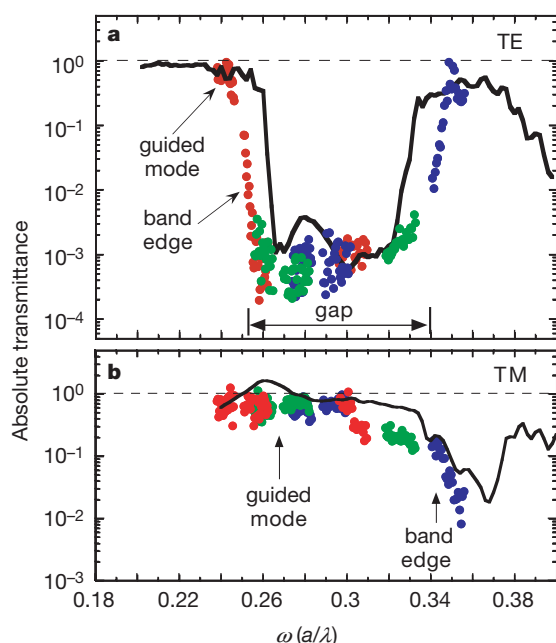


Figure 5 Absolute transmittance (T) as a function of $\omega (= a/\lambda)$ from 0.18 to 0.40. **a**, The TE transmittance spectrum covers the guided mode ($\sim 100\%$ transmission at $\omega = 0.24$), band edge, and bandgap regions. The solid line is a theoretical curve. Here, ω is varied by tuning λ through three samples with different a : 400 nm (red dots), 430 nm (green dots) and 460 nm (blue dots), respectively. **b**, The transmittance for TM input light. The solid line is a theoretical curve.

dominates and its intensity is close to that of total transmitted light. For $3 < N < 5$, the converted TM intensity increases, yet TE light continues to be reduced, leading to a large difference between TE and the total transmitted light. For $N > 5$, TM polarization dominates the total transmitted light and its intensity remains constant as it is a guided mode. At this point, a back-conversion from TM to TE becomes significant, leading to a nearly constant TE light intensity. The TE to TM ratio of about 0.05 for $N > 5$ is a measure of TE/TM mode conversion efficiency. This conversion process contributes to light leakage, and will limit the attenuating efficiency of a TE gap. Such leakage, however, may be eliminated through a symmetrical slab structure design.

Possible applications of this work in planar photonic circuits include low-loss in-plane waveguide bends²¹, waveguide crossings²², and photonic tunnelling²³ for wavelength division multiplexer/demultiplexer applications. We note that we have not obtained a complete 3D photonic bandgap²⁴ owing to the light cone, and some radiation losses are inevitable when the structure deviates from perfect periodicity—for example, when boundary effects, fabrication defects, and resonant cavities²³ are considered. Further work will therefore be required to minimize radiation losses in the above applications, aided by theoretical and numerical evidence that low-loss defects and cavities are possible in these structures¹⁷. □

Received 19 January; accepted 24 August 2000.

1. Yablonovitch, E. Photonic band-gap structures. *J. Opt. Soc. Am. B* **10**, 283–295 (1993).
2. John, S. Localization of light. *Phys. Today* **44**(3), 32–40 (May 1991).
3. Ho, K. et al. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Commun.* **89**, 413–416 (1994).
4. Ozbay, E. et al. Measurement of three-dimensional photonic band gap in a crystal structure made of dielectric rods. *Phys. Rev. B* **50**, 1945–1948 (1994).
5. Lin, S. Y. et al. A three-dimensional photonic crystal operating at infrared wavelengths. *Nature* **394**, 251–253 (1998).
6. Soukoulis, C. M. (ed.) *Photonic Band Gaps and Localization* (NATO ASI Series B, Physics Vol. 308, Plenum, New York, 1993).
7. Joannopoulos, J., Meade, R. & Winn, J. *Photonic Crystals* (Princeton, New York, 1995).
8. Robertson, W. M. et al. Measurement of photonic band structure in a two-dimensional periodic dielectric array. *Phys. Rev. Lett.* **68**, 2023–2026 (1992).

9. Lin, S. Y. et al. Investigation of absolute photonic band gaps in two-dimensional dielectric structures. *J. Mod. Opt.* **41**, 385–393 (1994).
10. Gruning, U. & Lehmann, V. Two-dimensional infrared photonic crystal based on microporous silicon. *Thin Solid Films* **276**, 151–154 (1996).
11. Rosenberg, A. et al. Near-infrared two-dimensional photonic band gap materials. *Opt. Lett.* **21**, 830–832 (1996).
12. Krauss, T. F. et al. Two-dimensional photonic band gap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
13. Labilloy, D. et al. Quantitative measurement of transmission, reflection, and diffraction of two-dimensional photonic band gap structures at near-infrared wavelengths. *Phys. Rev. Lett.* **79**, 4147–4150 (1997).
14. Benisty, H. et al. Optical and confinement properties of two-dimensional photonic crystals. *IEEE J. Light Wave Technol.* **17**, 2063–2077 (1999).
15. Kanskar, M. Observation of leaky slab modes in an air-bridged semiconductor waveguide with a two-dimensional photonic lattice. *Appl. Phys. Lett.* **70**, 1438–1440 (1997).
16. Villeneuve, P. R. et al. Three-dimensional photonic confinement in photonic crystals of low-dimensional periodicity. *IEE Proc. Optoelectron.* **145**, 384–390 (1998).
17. Johnson, S. G. et al. Guided modes in photonic crystal slabs. *Phys. Rev. B* **60**, 5751–5758 (1999).
18. Kish, F. A. Planar native-oxide index-guiding AlGaAs/GaAs quantum well heterostructure lasers. *Appl. Phys. Lett.* **59**, 1755–1757 (1991).
19. Kittel, C. *Introduction to Solid State Physics* Ch. 7, 185–189 (Wiley & Sons, New York, 1976).
20. Foresi, J. S. et al. Photonic bandgap micro-cavities in optical waveguides. *Nature* **390**, 143–145 (1997).
21. Lin, S. Y. et al. Experimental demonstration of guiding and bending of electromagnetic waves in a photonic crystal. *Science* **282**, 274–276 (1998).
22. Johnson, S. et al. Elimination of cross talk in waveguide intersections. *Opt. Lett.* **23**, 1855–1857 (1998).
23. Fan, S. et al. Channel drop tunneling through localized states. *Phys. Rev. Lett.* **80**, 960–963 (1998).
24. Fleming, J. G. & Lin, S. Y. Three-dimensional photonic crystal with a stop band from 1.35 to 1.95 μm . *Opt. Lett.* **24**, 49–51 (1999).

Acknowledgements

The work at Sandia National Laboratories was supported through the US DOE. Sandia is a multi-programme laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US DOE. The work at Massachusetts Institute of Technology was supported by MRSEC and by the NSF.

Correspondence and requests for materials should be addressed to S.Y.L. (e-mail: SLIN@sandia.gov).

Geometry-dominated fluid adsorption on sculpted solid substrates

C. Rascón & A. O. Parry

Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK

The shape^{1,2} and chemical composition³ of solid surfaces can be controlled at a mesoscopic scale. Exposing such structured substrates to a gas that is close to coexistence with its liquid phase can produce quite distinct adsorption characteristics compared to those of planar systems⁴, which may be important for technologies such as super-repellent surfaces^{5,6} or micro-fluidics^{7,8}. Recent studies have concentrated on the adsorption of liquids on rough^{9–11} and heterogeneous¹² substrates, and the characterization of nanoscopic liquid films¹³. But the fundamental effect of geometry on the adsorption of a fluid from the gas phase has hardly been addressed. Here we present a simple theoretical model which shows that varying the shape of the substrate can exert a profound influence on the adsorption isotherms of liquids. The model smoothly connects wetting and capillary condensation through a number of examples of fluid interfacial phenomena, and opens the possibility of tailoring the adsorption properties of solid substrates by sculpting their surface shape.

The interaction of a gas with a solid surface is usually mediated by an adsorbed (microscopically thin) layer of liquid which is responsible for the rich behaviour of the solid–gas interface¹⁴. When the gas coexists with its liquid, an increase in the temperature is accompanied by an increase in the thickness of this adsorbed layer which attains macroscopic character at the wetting tempera-

ture T_w . At this temperature, the contact angle of a macroscopic drop placed on the surface vanishes and the liquid, which is said to wet that solid surface, spreads. Alternatively, this situation can be reached for a fixed temperature T above T_w (but below the critical temperature T_c) by increasing the pressure towards the coexistence value $p_{co}(T)$ or, equivalently, by increasing the chemical potential towards $\mu_{co}(T)$ (see Fig. 1). This phase transition, known as complete wetting, has been intensely studied for fundamental as well as for important technological reasons¹⁴. Theoretical predictions confirmed by experiments have shown that the thickness l_π of the adsorbed layer diverges when it approaches coexistence with a non-universal exponent β_{co} which is dependent on the range of the particle interactions (solid–fluid and fluid–fluid) involved in the system: $l_\pi \approx |\Delta\mu|^{-\beta_{co}}$, where $\Delta\mu \equiv \mu - \mu_{co}$ is the chemical potential (relative to the chemical potential at coexistence)¹⁴ which, for dilute gases, is related to the partial pressure by $\Delta\mu = k_B T \log(p/p_{co}(T))$ where k_B is Boltzmann's constant. This prediction, however, is restricted to planar solid substrates.

One of the simplest geometrical configurations beyond the planar wall is a parallel slit geometry comprising two planar substrates a distance L' from each other. For this system, a different phenomenon occurs called capillary condensation^{15,16}. By increasing the chemical potential at a constant temperature $T > T_w$, the thickness of the adsorbed liquid layers on either wall grows according to a law similar to that for the planar system. However, when the chemical potential reaches a certain value $\Delta\mu^* < 0$, the space between the walls fills up with liquid, even though that phase is not thermodynamically stable in bulk. The location of this first-order phase transition is given, for sufficiently large values of L' , by the Kelvin equation

$$\Delta\mu^* = - \frac{2\sigma}{(\rho_L - \rho_G)L'} \quad (1)$$

where ρ_L and ρ_G are the densities of the liquid and the gas respectively and σ is the surface tension between these phases.

This simple comparison reveals the influence of geometry on adsorption isotherms: the continuous divergence dominated by the microscopic details of the interaction with the substrate (embodied in the exponent β_{co}), in the case of a planar wall, is transformed into a discontinuous jump at a value of the chemical potential $\Delta\mu^*$, which does not depend on microscopic properties of the substrate, in the case of the slit (see Fig. 1). Although the full phenomenology of wetting is understood with the help of effective interfacial models¹⁴, capillarity condensation is best described by density functional methods¹⁵.

Here we show, with a purely geometrical model, that there is a wealth of adsorption phenomena induced by different wall geometries, which smoothly interpolates between wetting and capillary condensation. The model, which is exact in the macroscopic limit, is not intended to be a microscopic description of these phenomena; instead, it should capture the essential features of geometrically dominated adsorption, which, owing to the lack of adequate models, remains largely unexplored.

In order to understand the connection between wetting and capillarity, we consider generalized solid wedges with cross-section described by a shape function $|x|^\gamma/L'^{\gamma-1}$ in the x direction (where L' is a length associated with the dimensions of the wall). The limiting cases, $\gamma = 0$ and $\gamma = \infty$, correspond to a planar substrate and a capillary slit respectively. Other intermediate cases include the linear ($\gamma = 1$) and parabolic ($\gamma = 2$) wedges as shown in Fig. 2. This system has been only partially studied: macroscopic^{17–19} and effective interfacial methods have been used to describe the adsorption isotherms at a linear wedge^{20–22} and at walls with $0 < \gamma \leq 1$ (in the asymptotic limit $\Delta\mu \rightarrow 0$)²³. Although for $\gamma > 1$ effective interfacial methods are problematic, they suggest a picture of isothermal adsorption dominated by geometry, thereby supporting the following geometrical approach. For a given temperature and chemical

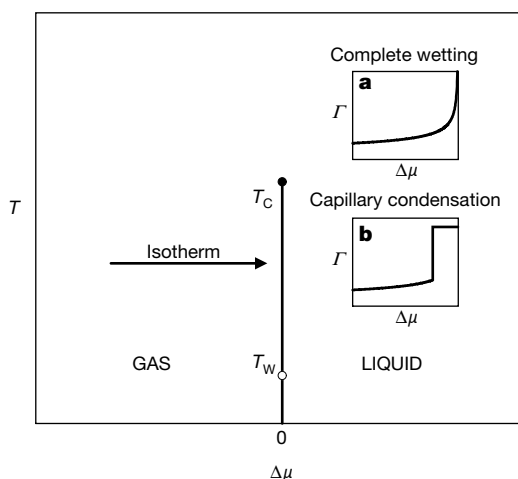


Figure 1 Schematic bulk phase diagram of a fluid showing the coexistence between gas and liquid phases. The gas phase is placed in contact with a solid substrate at a temperature between the critical temperature T_c and the wetting temperature T_w . As we approach coexistence, a liquid layer of increasing thickness is adsorbed on the substrate. Insets, adsorption isotherms are sketched for a planar substrate (**a**) and for a parallel slit (**b**).

potential, we can determine two pertinent lengths: the thickness of an adsorbed layer on a planar substrate, l_π , and the radius of curvature given by the Laplace equation $R \equiv \sigma/(\rho_L - \rho_G)|\Delta\mu|$. In order to construct the equilibrium profile of the adsorbed layer, we first coat the surface with a layer of thickness l_π (Fig. 3a and b). This is necessary to allow, in certain regimes, for the influence of microscopic forces (which are not accounted for in a purely macroscopic approach¹⁹). After that, a cylinder of radius R is fitted at the point of maximum curvature of the coated surface (Fig. 3c). The tangency points are the edges of a meniscus, whose shape is the lower part of the cylinder. The resulting interfacial shape can be determined by the combination of the coated surface and the meniscus (if any), and allow us to calculate the thickness l_0 of the adsorbed layer at the midpoint ($x = 0$) or the size of the meniscus as we approach coexistence (Fig. 3d).

This deceptively simple geometrical construction describes the gradual transformation of wetting (dominated by intermolecular forces) into capillary condensation (essentially dominated by geometry) as the substrate varies from being flat to being a parallel slit. The transformation is intricate and involves several asymptotic and pre-asymptotic regimes classified by the different behaviour of the midpoint thickness l_0 as a function of the chemical potential (Fig. 2). In our calculations, we have considered the experimentally relevant case of dispersion forces ($\beta_{co} = 1/3$). As we are interested in substrates with mesoscopic structure for which the influence of gravity is negligible, we have fixed the value of $A/12\pi\sigma L^2 \approx 10^{-3}$, where A is the Hamaker constant of the planar system²⁴. Considerably different values of this dimensionless quantity modify the presence of pre-asymptotic regimes but do not introduce any new feature into the phase diagram.

In the asymptotic limit ($\Delta\mu \rightarrow 0$), the continuous divergence of l_0 falls within one of three regimes, depending on the shape of the wall: a planar regime (PL) for $\gamma < \gamma^* \equiv 2\beta_{co}/(1 + \beta_{co})$, for which $l_0 \approx l_\pi \approx |\Delta\mu|^{-\beta_{co}}$; a first geometrically dominated regime (G_1) for $\gamma^* < \gamma < 1$ for which $l_0 \approx |\Delta\mu|^{-\gamma/(2-\gamma)}$; and a second geometrically dominated regime (G_2) for $1 < \gamma$ for which $l_0 \approx |\Delta\mu|^{-\gamma}$. We note that the exponent characterizing the divergence of l_0 varies continuously with the wedge-geometry exponent γ . The regimes PL and G_1 and the existence of a marginal case $\gamma = \gamma^*$ are in perfect agreement with the above prediction based on an effective hamiltonian model²³, thereby corroborating the strong

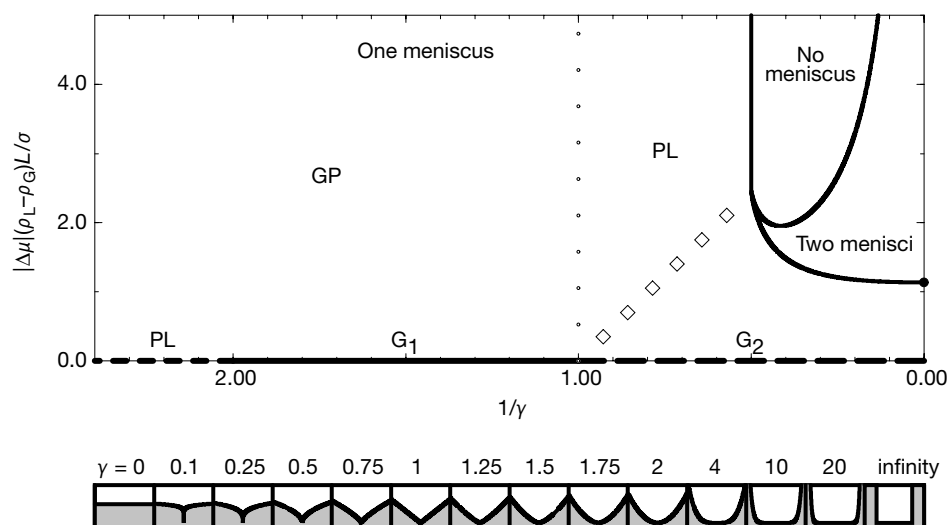


Figure 2 Phase diagram for the adsorption isotherms at generalized solid wedges with cross-section proportional to $|x|^\gamma$ in the x direction. Diagram was obtained using the geometrical construction of Fig. 3. The thickness (l_0) of the adsorbed layer at the mid-point ($x = 0$) grows according to different power laws as a function of the chemical potential $\Delta\mu$. We have chosen the experimentally relevant case of dispersion forces ($\beta_0 = 1/3 \Rightarrow \gamma^* = 1/2$) and fixed the dimensionless variable $A/12\pi\sigma L^2 \approx 10^{-3}$, where A is the Hamaker constant of the planar system. Approaching the bulk liquid–vapour coexistence, there are three asymptotic regimes: planar (PL) for

$\gamma < \gamma^* \equiv 2\beta_{co}/(1 + \beta_{co})$, and two geometrically dominated regimes G_1 and G_2 . Furthermore, different pre-asymptotic regimes appear for larger values of $|\Delta\mu|$: for $\gamma < 1$, a regime combining planar and geometric features (GP), a planar regime for $1 < \gamma < 2$, a regime in which no meniscus is formed, and regime with two symmetrically placed menisci. In the limit $\gamma \rightarrow \infty$, the transition between the two-menisci regime and the central-meniscus regime gives rise to capillary condensation (filled dot). The lower diagram shows the evolution of the cross-section with the parameter γ ranging from a planar wall ($\gamma = 0$) to a capillary slit ($\gamma = \infty$).

influence of geometry on liquid adsorption. The regime G_2 , to our knowledge not previously described in the literature, can be employed to fabricate microfluidic devices with an arbitrarily sensitive response to changes in partial pressure.

In addition, several pre-asymptotic regimes are also captured by this description. Figure 4 shows the typical evolution of the mid-point thickness l_0 as a function of the chemical potential for different wall shapes. For $\gamma = 0.25$ ($\gamma < \gamma^*$), two distinct regimes appear (Fig. 4a). Together with the planar asymptotic regime close to coexistence, a regime (GP) involving geometric and planar features appears for which the mid-point thickness grows like $l_0 \approx l_\pi^*$. This regime persists for $\gamma^* < \gamma < 1$, as can be seen in Fig. 4b for $\gamma = 0.75$, although the asymptotic behaviour has changed and is geometrically driven. In contrast, for $\gamma > 1$, along

with a unique asymptotic behaviour, there are two distinct pre-asymptotic regimes separated by the marginal case $\gamma = 2$. For $1 < \gamma < 2$, l_0 grows in a planar-like manner far from coexistence and presents a crossover to the true asymptotic regime G_2 , as can be seen in Fig. 4c for $\gamma = 1.5$. Within our approximation, the chemical potential always enters as the dimensionless variable $\Delta\mu(\rho_L - \rho_G)L/\sigma$. Depending on the value of each of these quantities, the experimental window might reach only one of the regimes described above.

The singular case $\gamma = 2$, the parabolic wedge, deserves special attention as it gives rise to a new phenomenon. For this case, no meniscus forms until the chemical potential has exceeded a certain threshold. This emerges as a direct consequence of the geometry. If the radius of curvature given by Laplace's equation is much smaller than the minimum radius of curvature of the surface ($R \ll L/2$), no meniscus is developed and a thin layer covers the substrate. The value of the chemical potential for which the meniscus starts to form is given implicitly, in this simple geometrical approximation, by $R + l_\pi = L/2$, where l_π represents a correction due to the thickness of the adsorbed layer. Once the meniscus has originated, the mid-point thickness tends to grow as dictated by the asymptotic limit G_2 . We refer to this phenomenon as the meniscus transition although fluctuation effects, not described here, transform it into a smooth, but nonetheless abrupt, crossover from microscopic to macroscopic adsorption, characterized by scaling behaviour. It can be considered a legitimate phase transition in the macroscopic limit ($l_\pi/L \rightarrow 0$) for which the present geometrical model is exact.

A new phenomenon, leading to capillary condensation in the limit $\gamma \rightarrow \infty$, appears for $\gamma > 2$. For these substrates, the point of minimum curvature has shifted from $x = 0$ and, owing to the symmetry $x \leftrightarrow -x$ two menisci materialize in two (symmetric) points of the wall at a certain value of the chemical potential. Because those two points are locally parabolic, this effect is the same as the meniscus transition described for the case $\gamma = 2$. However, by approaching coexistence more closely, these menisci increase in size and eventually merge into a unique central meniscus. This

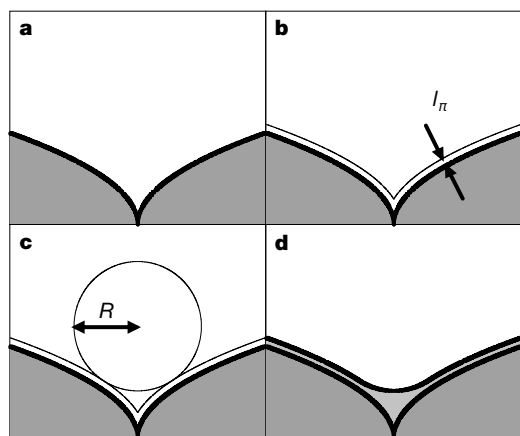


Figure 3 Geometrical construction for determining the shape of the adsorbed layer. The surface (a) is first coated with a thin film of thickness l_π , the thickness of an adsorbed layer on a planar substrate (b). Then, a cylinder of radius R , given by Laplace's equation, is fitted in the point of minimum curvature (c). The final shape is given by the continuous combination of both shapes (d).

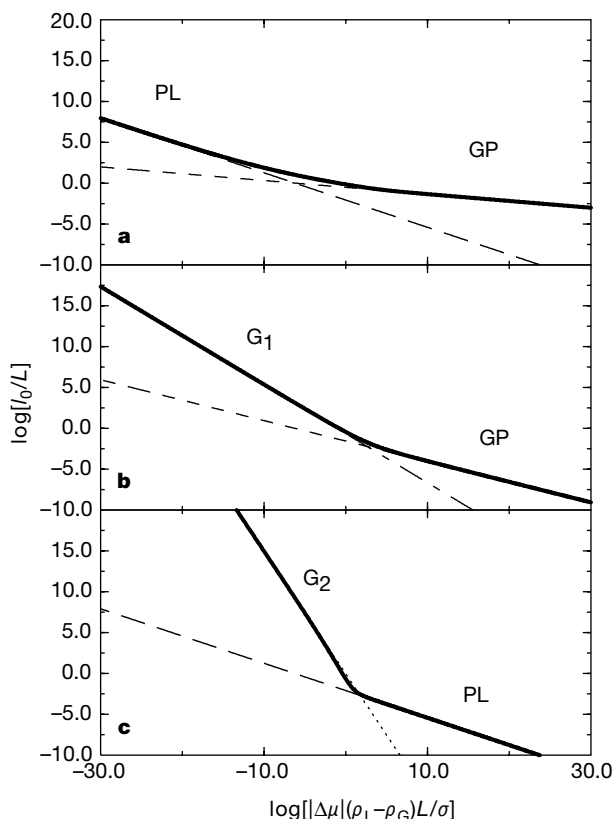


Figure 4 Adsorption isotherms for solid wedges with cross-section described by a shape function proportional to $|x|^\gamma$ in the x direction. The thickness of the adsorbed layer at the midpoint $x = 0$, calculated with the geometrical method described in the text, is represented as a function of the chemical potential $\Delta\mu$ for (a) $\gamma = 0.25$, (b) $\gamma = 0.75$ and (c) $\gamma = 1.5$. The scale is the same for all the wall shapes in order to facilitate the comparison. The pure asymptotic regimes correspond to the straight lines: PL (long dashed), GP (short dashed), G_1 (dot-dashed) and G_2 (dotted).

profoundly affects the behaviour of the system. Once this unique meniscus is developed, the thickness at the midpoint is extremely sensitive to changes in the chemical potential and approximates the asymptotic limit $l_0 \approx |\Delta\mu|^{-\gamma}$. In the limiting case $\gamma = \infty$, the merging of the two menisci precipitates the immediate and abrupt rise of the liquid level and the space between the walls fills up completely, giving rise to capillary condensation. As the parallel walls are connected by the bottom of the wedge, no metastable states are expected²⁵. The geometrical description of this phenomenon predicts that the transitions take place at a value of the chemical potential given by $R + l_\pi = L$, which recovers the Kelvin equation (1) with a correction, given by l_π , that is exact for systems with short-ranged intermolecular forces¹⁵.

This extraordinarily rich sequence of different regimes and phenomena induced by the substrate shape are shown in the phase diagram of Fig. 2. A similar phase diagram is expected for radially symmetric substrates described by a shape function proportional to r^γ , which ranges between a flat surface and a capillary tube. These distinct substrate shapes could be used as microfluidic devices to place mesoscopic amounts of liquid at preferential points of the surface of a solid substrate in a controlled manner by varying only the partial pressure. □

Received 18 April; accepted 29 August 2000.

- Xia, Y. & Whitesides, G. M. Soft lithography. *Angew. Chem. Int. Edn* **37**, 550–575 (1998).
- Trau, M. *et al.* Microscopic patterning of orientated mesoscopic silica through guided growth. *Nature* **390**, 674–676 (1997).
- Kumar, A., Abbott, N. A., Kim, E., Biebuyck, H. A. & Whitesides, G. M. Patterned self-assembled monolayers and meso-scale phenomena. *Acc. Chem. Res.* **28**, 219–226 (1995).

- Dietrich, S. in *Proceedings of the NATO-ASI "New Approaches to Old and New Problems in Liquid State Theory"* (eds Caccamo, C., Hansen, J. P. & Stell, G.) 197–244 (Kluwer, Dordrecht, 1999).
- Shibuchi, S., Yamamoto, T., Onda, T. & Tsujii, K. Super water- and oil-repellent surfaces resulting from fractal structure. *J. Colloid Interf. Sci.* **208**, 287–294 (1998).
- Bico, J., Marzolin, C. & Quere, D. Pearl drops. *Europhys. Lett.* **47**, 220–226 (1999).
- Weigl, B. H. & Yager, P. Microfluidics—Microfluidic diffusion-based separation and detection. *Science* **283**, 346–347 (1999).
- Gravens, P., Branebjerg, J. & Jensen, O. S. Microfluidics—a review. *J. Micromech. Microeng.* **3**, 168–182 (1993).
- Kardar, M. & Indekeu, J. O. Adsorption and wetting transitions on rough substrates. *Europhys. Lett.* **12**, 161–166 (1990).
- Giugliarelli, G. & Stella, A. L. Discontinuous interface depinning from a rough wall. *Phys. Rev. E* **53**, 5035–5038 (1996).
- Netz, R. R. & Andelman, D. Roughness-induced wetting. *Phys. Rev. E* **55**, 687–700 (1997).
- Gau, H., Herminghaus, S., Lenz, P. & Lipowsky, R. Liquid morphologies on structured surfaces: From microchannels to microchips. *Science* **283**, 46–49 (1999).
- Luna, M., Colchero, J. & Baro, A. M. Study of water droplets and films on graphite by noncontact scanning force microscopy. *J. Phys. Chem. B* **103**, 9576–9581 (1999).
- Dietrich, S. in *Phase Transitions and Critical Phenomena* (eds Domb, C. & Lebowitz, J. L.) Vol. 12, 1–218 (Academic, London, 1988).
- Evans, R., Marconi, U. M. B. & Tarazona, P. Fluids in narrow pores: Adsorption, capillary condensation and critical points. *J. Chem. Phys.* **84**, 2376–2399 (1986).
- Christenson, H. K. Capillary condensation due to van der Waals attraction in wet slits. *Phys. Rev. Lett.* **73**, 1821–1824 (1994).
- Concus, P. & Finn, R. On the behaviour of a capillary surface in a wedge. *Proc. Natl Acad. Sci. USA* **63**, 292–299 (1969).
- Pomeau, Y. Wetting in a corner and related questions. *J. Colloid. Interf. Sci.* **113**, 5–11 (1986).
- Hauge, E. H. Macroscopic theory of wetting in a wedge. *Phys. Rev. A* **46**, 4994–4998 (1992).
- Neimark, A. V. & Kheifets, L. I. Solution of the problem of the equilibrium profile of the transition zone between a wetting film and the meniscus of the bulk phase in capillaries. *Colloid. J. USSR* **43**, 402–407 (1981).
- Rejmer, K., Dietrich, S. & Napiórkowski, M. Filling transition for a wedge. *Phys. Rev. E* **60**, 4027–4042 (1999).
- Parry, A. O., Rascón, C. & Wood, A. J. Critical effects at 3D wedge wetting. *Phys. Rev. Lett.* **85**, 345–348 (2000).
- Rascón, C. & Parry, A. O. Geometry dependent critical exponents at complete wetting. *J. Chem. Phys.* **112**, 5175–5180 (2000).
- Israelachvili, J. *Intermolecular & Surface Forces* (Academic, London, 1991).
- Marconi, U. M. B. & Van Swol, F. Microscopic model for hysteresis and phase equilibria of fluids confined between parallel plates. *Phys. Rev. A* **39**, 4109–4116 (1989).

Acknowledgements

C.R. acknowledges financial support from the European Commission.

Correspondence and requests for materials should be addressed to A.O.P. (e-mail: a.o.parry@ic.ac.uk).

Influence of mean climate change on climate variability from a 155-year tropical Pacific coral record

Frank E. Urban*, Julia E. Cole†‡ & Jonathan T. Overpeck*‡

* Department of Geological Sciences/INSTAAR, Box 450, University of Colorado, Boulder, Colorado 80309, USA

† Department of Geological Sciences/INSTAAR/PAOS, Box 450, University of Colorado, Boulder, Colorado 80309, USA

Today, the El Niño/Southern Oscillation (ENSO) system is the primary driver of interannual variability in global climate, but its long-term behaviour is poorly understood. Instrumental observations reveal a shift in 1976 towards warmer and wetter conditions in the tropical Pacific, with widespread climatic and ecological consequences^{1–3}. This shift, unique over the past century⁴, has prompted debate over the influence of increasing atmospheric

‡ Present addresses: Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA (J.E.C.); Institute for the Study of Planet Earth/Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA (J.T.O.).

concentrations of greenhouse gases on ENSO variability^{5–7}. Here we present a 155-year ENSO reconstruction from a central tropical Pacific coral that provides new evidence for long-term changes in the regional mean climate and its variability. A gradual transition in the early twentieth century and the abrupt change in 1976, both towards warmer and wetter conditions, co-occur with changes in variability. In the mid–late nineteenth century, cooler and drier background conditions coincided with prominent decadal variability; in the early twentieth century, shorter-period (~2.9 years) variability intensified. After 1920, variability weakens and becomes focused at interannual timescales; with the shift in 1976, variability with a period of about 4 years becomes prominent. Our results suggest that variability in the tropical Pacific is linked to the region's mean climate, and that changes in both have occurred during periods of natural as well as anthropogenic climate forcing.

Before about 1950, spatial coverage of instrumental climate data in the central tropical Pacific diminishes sharply⁸. Longer instrumental ENSO indices are often based on interpolation among increasingly sparse observations⁹ or on single sites that may not consistently represent ENSO centres of action¹⁰ or where data quality degrades with time⁶. Palaeoclimatic records reveal longer term variability^{4,11–16}. However, existing reconstructions are qualitative, are based upon ENSO's remote impacts in the extratropics, do not substantially extend the instrumental record, or have chronological uncertainties that complicate the reconstruction of annual anomalies. Coral records from the western Indian Ocean suggest that decadal variability in this region may originate from the tropical Pacific¹⁷, but sufficiently long records from the tropical Pacific have not been available for comparison. Here we present a 155-year coral record from the central tropical Pacific that provides new evidence of long-term Pacific climate variability, including substantial decadal variability in the mid to late nineteenth century and a trend towards warmer, wetter conditions from 1840 to 1995.

The ENSO signal in coral $\delta^{18}\text{O}$ records from the central western Pacific is strong and unambiguous^{4,18,19}. In this region (Fig. 1), El Niño events bring small sea surface temperature (SST) increases of less than 1.5 °C and enhanced rainfall associated with the migration of intense convection from the Indonesian region, as the warm, fresh pool of the western Pacific expands eastward²⁰. These changes combine to generate negative $\delta^{18}\text{O}$ anomalies in the coral skeleton¹⁸. La Niña events experience relatively cool and dry conditions that produce positive coral $\delta^{18}\text{O}$ anomalies.

The bimonthly $\delta^{18}\text{O}$ record from Maiana (Fig. 2, top panel) reveals new aspects of long-term variability in this region. A trend towards warmer, wetter conditions spans the full record; this trend occurs in two stages, gradually in the early twentieth century, and more abruptly in 1976, with relatively stationary conditions elsewhere. The early record (1840–95) clearly shows decadal variability,

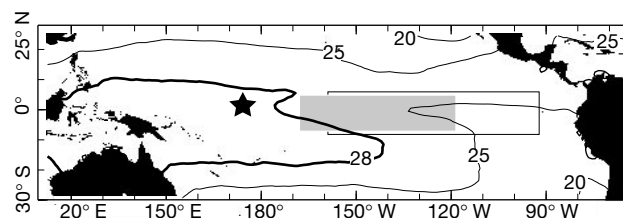


Figure 1 Map of tropical Pacific, showing locations of instrumental and coral records used. Maiana Atoll is indicated by a star, and the warm pool is indicated by the 28 °C isotherm. Also shown are the regions used to define large-scale ENSO indices: the Niño 3.4 SST region is outlined by a box, and the region used for the central Pacific rainfall index is shaded. The Multivariate ENSO Index includes information on surface air and ocean temperatures, wind, cloudiness and sea level pressure from the full Pacific Ocean region shown here.

in contrast to the mainly interannual variability that characterizes the twentieth century. This decadal variability is strongest when the background conditions are cooler and drier (mid to late nineteenth century). Finally, consistent with previous work^{4,21}, interannual variance of the twentieth century is attenuated between 1920 and 1950. A detailed comparison of the multivariate ENSO index (MEI)²² and the Niño 3.4 sea surface temperature (SST)⁹ since 1950 shows that the Maiana record corresponds closely to the state of ENSO (Fig. 2, bottom panel).

The long-term trend in the Maiana record spans a range of approximately 0.55‰ $\delta^{18}\text{O}$, with one-third of this change occurring in the early twentieth century and the remaining two-thirds in the 1976 shift. Interpreted strictly as temperature, the full trend would represent a warming of 2.3–3.1 °C, using a range of published temperature– $\delta^{18}\text{O}$ relationships for this coral genus (0.18–0.24‰ per 1 °C)^{17,23}. Interpreted strictly as salinity, this change would represent a 2.0‰ freshening, based on an average relationship in the central western Pacific of 0.27‰ $\delta^{18}\text{O}$ per 1‰ salinity⁸. This salinity– $\delta^{18}\text{O}$ relationship implies a rainwater isotopic value of about –9‰ (ref. 8). Complicating this linear interpretation, we suggest that past cooler and drier periods may have experienced less intense rainfall events; thus local rainwater would have less depleted $\delta^{18}\text{O}$ values than today, amplifying the $\delta^{18}\text{O}$ difference from modern.

Modern climate relationships in the equatorial Pacific indicate that warming should be accompanied by freshening, either through intensified rainfall over warmer waters or through eastward expansion of the western Pacific warm and fresh pool²⁰. Thus the trend in coral $\delta^{18}\text{O}$ probably reflects a combination of warming and freshening. Whether it occurs by advection or by local warming and rainfall, this change represents an expansion of the western Pacific warm pool over the past 155 years, as Maiana lies at the eastern edge

Table 1 Linear zero-lag correlations among bimonthly coral and climate records, 1950–89

	Mai 1-1	Ara 2-1	Tar 89	Average	MEI	Niño 3.4	CP Rain	DSLIP	TSLP	SOI
1. Mai 2-3	0.78	0.70	0.67	0.88	–0.76	–0.58	–0.68	–0.50	0.40	0.53
2. Mai 1-1		0.86	0.66	0.93	–0.75	–0.65	–0.82	–0.58	0.46	0.61
3. Ara 2-1			0.66	0.91	–0.75	–0.66	–0.79	–0.62	0.46	0.63
4. Tar 89				0.84	–0.71	–0.59	–0.66	–0.51	0.36	0.49
5. 4-coral average					–0.82	–0.70	–0.83	–0.62	0.47	0.63
6. MEI						0.89	0.84	0.80	–0.67	–0.86
7. Niño 3.4							0.77	0.73	–0.56	–0.73
8. CP rain								0.67	–0.51	–0.67
9. DSLIP									–0.46	–0.84
10. TSLP										0.84

Abbreviations and sources: 1, coral data (this work); 2 and 3, data from J.E.C. *et al.* (manuscript in preparation); 4, Tarawa coral data; 5, average of four previous coral records; MEI, Multivariate ENSO index²²; 7, Niño 3.4 SST⁹; CP rain, Central Pacific rainfall (available at <http://tao.atmos.washington.edu/datasets/eqpacislandrainindex/>); DSLIP, Darwin sea level pressure anomaly; TSLP, Tahiti sea level pressure anomaly; SOI, Southern Oscillation Index²⁰ (updated by P. Jones, personal communication). All correlations are significant at more than 99%. Abbreviations used: Mai 2-3, Maiana 95-2-3; Mai 1-1, Maiana 95-1-1; Ara 2-1, Aranuka 95-2-1; Tar 89, Tarawa 89.

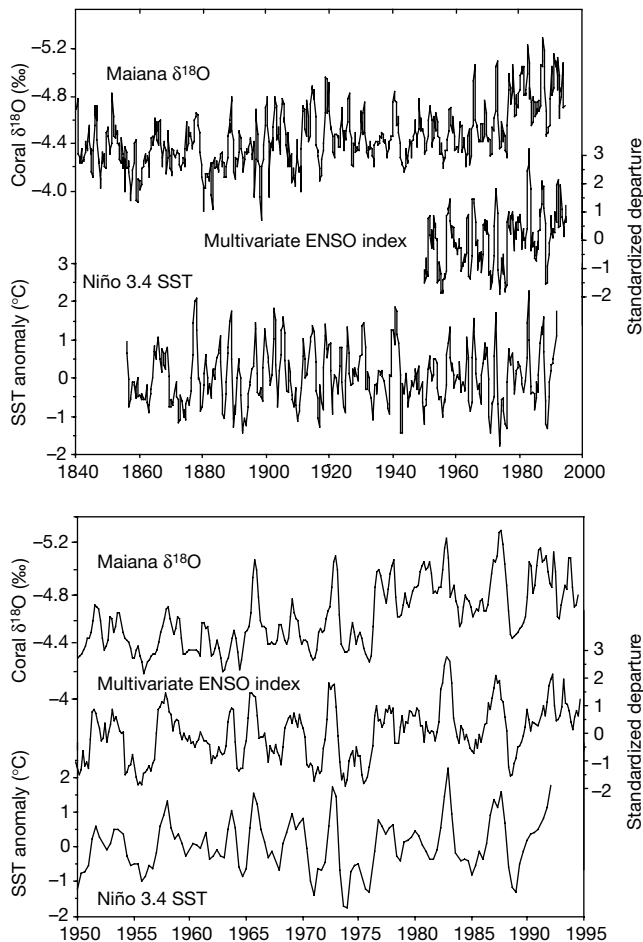


Figure 2 Bimonthly records of tropical Pacific variability from coral and instrumental data. Top panel, Maiana coral $\delta^{18}\text{O}$ compared to the Multivariate ENSO index and Niño 3.4 SST. Bottom panel, expansion of these records over the 1950–95 period shows detailed correlation among these records.

of the existing warm pool (Fig. 1) in a region of strong zonal temperature and salinity gradients.

Most of the Maiana coral $\delta^{18}\text{O}$ trend ($\sim 0.35\text{‰}$ $\delta^{18}\text{O}$) occurred in the 1976 shift. Instrumental records from the COADS data set²⁴ indicate a warming at Maiana of 0.6°C relative to the mean of the previous 25 years. This change is about twice the warming seen in the Niño 3.4 region further east and is equivalent to a decrease of $\sim 0.13\text{‰}$ coral $\delta^{18}\text{O}$. The coral record thus implies a freshening of 0.8‰ salinity (associated with the additional decrease of 0.22‰ $\delta^{18}\text{O}$) at this time. Increased rainfall⁵ accounts for at least part of this change, with eastward advection of fresher water from the warm pool probably making up the balance²⁰. Instrumental salinity data over the 1976 transition are insufficient to evaluate these inferences, but salinity at the Tropical Atmosphere Ocean (TAO; www.pmel.noaa.gov/toga-tao/home.html) array mooring at 0° , 165°E drops by about 1‰ with the onset of the prolonged El Niño of the early 1990s, suggesting that the inferred magnitude of change is reasonable.

To characterize our results in the frequency domain, we applied spectral and cross-spectral analyses using a multitaper method with red noise assumptions²⁵. Spectral analysis of the full record (1840–1995; not shown) reveals significant ($>90\%$) peaks centred approximately at periods of 10–15, 4.7–6, 3.3–3.8, 2.5–2.8 and 2.2–2.4 years. Cross-spectral analysis between the Maiana MAI95–2–3 record and Niño 3.4 SST (1856–1992)⁹ reveals coherency above the 95% significance level across nearly all periods from 2–12 years

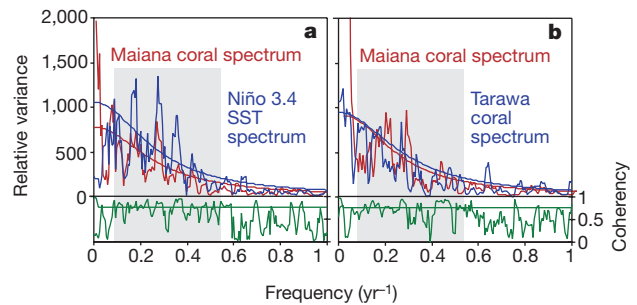


Figure 3 Results of cross-spectral analysis²⁵ between the normalized Maiana coral record and other normalized Pacific datasets. The Maiana variance (in red) is compared with **a**, Niño 3.4 SST (1856–1992; in blue) and **b**, a coral record from Tarawa Atoll (1893–1989; in blue). Smooth red and blue lines are 90% confidence limits. Coherency between records is shown in green, with the 95% confidence level indicated. The Maiana record is coherent ($>95\%$ confidence) with both the Niño 3.4 and the Tarawa records for all significant concentrations of variance at periods of 2–12 years (shaded). Maiana variance spectra differ between these panels because the interval of analysis differs; the univariate spectrum of the full-length Maiana time series (not shown) has significant features (see text).

(Fig. 3a); the decadal variance in Niño 3.4 is clearly captured by the coral record. A cross-spectral comparison of the Maiana and Tarawa coral data⁴ (1893–1989) display similar coherent concentrations of variance (Fig. 3b).

Evolutionary multitaper spectral analysis²⁵ of Maiana coral and Niño 3.4 SST records reveals a highly variable spectrum of ENSO variability since AD 1840 (Fig. 4). The input data consist of 40-year subsets of bimonthly data that are overlapped in 4-year steps; variance concentrations are plotted as a function of frequency at the centrepoint year for each interval. Before 1890, the coral record shows persistent decadal variability centred at 12.5 years; the Niño 3.4 spectrum hints at this feature but is barely long enough to reach this interval. At this time, interannual variance is muted in the coral record. Around 1900 a transition begins, characterized by a shift of variance from decadal toward shorter periods (5–7 years). Variance centred on 2.9 years emerges around 1880 and persists until about 1920. At this point, the 2.9-year variance disappears and interannual variance weakens generally, except for weak persistence at the 5–7-year period. Around 1955, the variance shifts towards broad concentration centred at an approximately 4-year period, which persists and intensifies through the end of the record. The low-frequency variance concentration in the final years of the coral record reflects the step-function change in 1976, which is more pronounced in the Maiana record than in the Niño 3.4 index.

Previous work has attributed differences between the recent spectral signature of ENSO (1960–1998) and the average spectral signature (1856–1976) to the influence of anthropogenic greenhouse forcing on post-1976 variability²⁶. However, our results confirm^{4,13–15} that the spectrum of ENSO variability changes substantially over multidecadal timescales before strong anthropogenic forcing (see the pre-1950s variability in Fig. 4, for example). Over the period of the Maiana record, there is no fixed natural spectrum of ENSO variability, and the variance spectrum of the post-1976 Maiana record is not unique. However, our record shows that the warming and freshening in this region since 1976 is unprecedented since AD 1840.

The changes in dominant periods of variance in the Maiana coral record occur with the changes in the background mean $\delta^{18}\text{O}$, suggesting that the frequency-domain characteristics of climate variability at this site are sensitive to background conditions. Decadal variability is strongest, and interannual variability weakest, when background conditions are relatively cooler and drier (in the

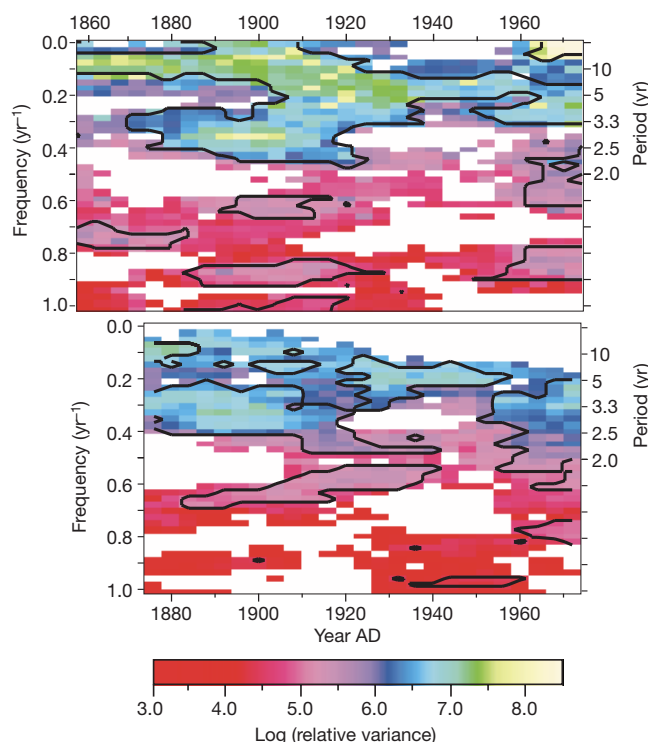


Figure 4 Evolutionary spectral analysis results. Top, Maiana; bottom, Niño 3.4 SST records. Analysis was performed on 40-year segments of data overlapped by four years, using multitaper methods with red noise background assumptions²⁵. Coloured regions show variance significant above the median background (>50%), and black contours enclose regions with variance significant at >90%. These spectra share several general features:

mid to late nineteenth century). This may explain the attenuated interannual variability during this time that is implied by a global network of palaeoclimate records largely outside the ENSO centre of action²⁷. The 2.9-year component coincides with a background transition in the early twentieth century; the second transition (in 1976) may also be associated with the resurgence of variance at a 4-year period.

The changes in time and frequency domains at Maiana do not correlate linearly with global radiative forcings (solar, volcanic or greenhouse-gas concentrations). However, the Maiana record shift to warmer and wetter conditions in 1976 coincides with the shift to warmer conditions that appears clearly in hemispheric and global temperature indices²⁸. This change has been attributed to the intensification of the tropical hydrological cycle by anthropogenic greenhouse forcing⁵, which has a strong regional signature in the central equatorial Pacific consistent with the substantial $\delta^{18}\text{O}$ shift in our record.

The intensified decadal variability in the early part of the Maiana record represents a fundamental change in the behaviour of ENSO, but additional records from the central tropical Pacific are needed to assess the nature of that change. One possibility is that the timescale of ENSO has shifted from decadal to interannual since the mid to late nineteenth century. If true, this shift may require rethinking standard models of ENSO, which have a characteristic timescale of 3–5 years. Alternatively, ENSO may have changed its spatial signature such that regions of the central tropical Pacific that are currently the focus of interannual ENSO impact (particularly tropical convection anomalies) were less strongly affected by this variability in the mid to late nineteenth century. Past climate may then have imposed a mix of temperature and salinity components into the $\delta^{18}\text{O}$ signal different from the inferred salinity-dominated signal of the modern record. A third possibility, that the coral record

significant nineteenth-century decadal variance, strong variance at 2.9-year periods in the early twentieth century, attenuation of 2–4-year variance between 1920–55, and the general progression from decadal towards higher-frequency interannual variance over the length of the record. The low-frequency variance concentration in the final years of the coral record reflects the step-function change in 1976.

reflects non-climatic or biological variability, seems unlikely given the high correlation and decadal coherency among long coral and instrumental records.

Decadal climate variability in the tropical Pacific is the subject of intense continuing study, largely focused on the shift in 1976. Our data extend the observational window into the pre-industrial era and provides new evidence that decadal variance was stronger in the tropical Pacific in the mid to late 1800s, when background conditions were cooler and drier and anthropogenic greenhouse forcing was absent. Incorporating the physical mechanisms responsible for this timescale of variability will be an important challenge for predictive ENSO models; reconstructing the spatial fingerprint and extratropical impacts of this variability will require greater attention to development and analysis of pre-twentieth century climate records. □

Methods

Sampling and isotopic analysis

We have produced a new 155-year $\delta^{18}\text{O}$ record from a *Porites* spp. coral that grew on the outer reef at Maiana Atoll, Republic of Kiribati (Fig. 1; 1° N, 173° E). Cores (8.5 cm in diameter) drilled from living coral colonies were slabbed and X-rayed to identify annual density bands and optimal sampling transects. Samples were drilled along the growth axis continuously at 1mm resolution. Stable isotopic analysis ($\delta^{18}\text{O}$ and $\delta^{13}\text{C}$) of powdered samples was performed on a Micromass Optima with Isocarb automated preparation system at the University of Colorado's Institute for Arctic and Alpine Research. Analytical precision is ± 0.08 ‰ $\delta^{18}\text{O}$ and ± 0.05 ‰ $\delta^{13}\text{C}$. The Maiana record is continuous except for a gap from 1901–05, which we bridged through straightforward comparison and splicing of data from the Tarawa coral record⁴.

Age model

Seasonally varying $\delta^{13}\text{C}$ values and ephemeral annual density bands provide the basis for annual age assignment. Ages were adjusted subseasonally by comparison to major El Niño and La Niña anomalies in a regional rainfall index (available at <http://tao.atmos>).

washington.edu/datasets/eqpacislandrainindex/); this procedure allows us to align major peaks and troughs that occur within the same calendar year with adjustments that are generally less than 3 months. Because corals grow at variable rates throughout the year²⁹, these small chronological adjustments are reasonable and customary (and do not affect estimates of multiyear variability). Near the base of the core, an age adjustment of 1 year was needed, suggesting an age uncertainty on the full record of ± 1 year. The Maiana coral grew at a rate of 7–12 mm per year; we interpolated records to a common time step of bimonthly resolution for further analysis.

Replication

The record from core MAI95-2-3 is well replicated by additional records of 65–70 years from Maiana and Aranuka, about 75 km southeast of Maiana (J.E.C. *et al.*, manuscript in preparation) and by a published 97-year record from Tarawa Atoll, about 30 km north of Maiana⁴. Correlation coefficients among coral and instrumental data over the common period 1950–89 are consistently high (Table 1), adding confidence to single-core reconstruction of regional variability in western Kiribati. We focus on the MAI95-2-3 record for further analysis because it extends the record of ENSO into pre-instrumental periods and avoids statistical nonstationarity that may be introduced through inclusion of records that differ in length. The 4-coral average correlates slightly better with instrumental data than most individual coral records, but only the MAI95-2-3 core extends the instrumental record of ENSO. Linear correlation coefficients demonstrate a strong agreement between the Maiana record and ENSO variability (Table 1), and decadal averages of the Maiana record correlate significantly over the past century (>95% confidence) with those from the Tarawa coral record and regional rainfall indices, substantiating the reality of these more subtle low-frequency signals (J.E.C. *et al.*, manuscript in preparation).

Received 1 February; accepted 9 August 2000.

- Zhang, Y., Wallace, J. M. & Battisti, D. S. ENSO-like interdecadal variability: 1900–93. *J. Clim.* **10**, 1004–1020 (1997).
- Trenberth, K. E. & Hurrell, J. W. Decadal atmosphere-ocean variations in the Pacific. *Clim. Dyn.* **9**, 303–319 (1994).
- Ebbesmeyer, C. C. *et al.* in *Proc. 7th Annu. Pacific Climate (PACCLIM) Workshop* (eds Betancourt, J. L. & Sharp, V. L.) 120–141 (Interagency Ecological Studies Program Technical Report No. 26, California Dept of Water Resources, Sacramento, 1991).
- Cole, J. E., Fairbanks, R. G. & Shen, G. T. The spectrum of recent variability in the Southern Oscillation: results from a Tarawa Atoll coral. *Science* **262**, 1790–1793 (1993).
- Graham, N. E. Simulation of recent global temperature trends. *Science* **267**, 686–671 (1995).
- Trenberth, K. E. & Hoar, T. W. The 1990–1995 El Niño–Southern Oscillation event: Longest on record. *Geophys. Res. Lett.* **23**, 57–60 (1996).
- Rajagopalan, B., Lall, U. & Cane, M. A. Anomalous ENSO occurrences: an alternate view. *J. Clim.* **10**, 2351–2357 (1997).
- Fairbanks, R. G. *et al.* Evaluating climate indices and their geochemical proxies measured in corals. *Coral Reefs* **16**, S93–S100 (1997).
- Kaplan, A. *et al.* Analyses of global sea surface temperature, 1856–1991. *J. Geophys. Res.* **103**, 18567–18589 (1998).
- Allan, R., Lindesay, J. & Parker, D. *El Niño Southern Oscillation and Climatic Variability* (CSIRO, Collingwood, 1997).
- Quinn, W. H., Neal, V. T. & Antuñez de Mayolo, S. E. El Niño occurrences over the past four and a half centuries. *J. Geophys. Res.* **92**, 14449–14461 (1987).
- Kennedy, J. A. & Brasseur, S. C. Molecular records of twentieth-century El Niño events in laminated sediments from the Santa Barbara basin. *Nature* **357**, 62–64 (1992).
- Dunbar, R. B., Wellington, G. M., Colgan, M. W. & Glynn, P. W. Eastern Pacific sea surface temperature since 1600 AD: The $\delta^{18}\text{O}$ record of climate variability in Galapagos corals. *Paleoceanography* **9**, 291–316 (1994).
- Tudhope, A. W. *et al.* Recent changes in climate in the far western equatorial Pacific and their relationship to the Southern Oscillation: oxygen isotope records from massive corals, Papua New Guinea. *Earth Planet. Sci. Lett.* **136**, 575–590 (1995).
- Stahle, D. W. *et al.* Experimental dendroclimatic reconstruction of the Southern Oscillation. *Bull. Am. Meteorol. Soc.* **79**, 2137–2152 (1998).
- Evans, M. N., Fairbanks, R. G. & Rubenstone, J. L. A proxy index of ENSO teleconnections. *Nature* **394**, 732–733 (1998).
- Cole, J. E., Dunbar, R. B., McClanahan, T. R. & Muthiga, N. Tropical Pacific forcing of decadal variability in the western Indian Ocean over the past two centuries. *Science* **287**, 617–619 (2000).
- Cole, J. E. & Fairbanks, R. G. The Southern Oscillation recorded in the oxygen isotopes of corals from Tarawa Atoll. *Paleoceanography* **5**, 669–683 (1990).
- Guilderson, T. P. & Schrag, D. P. Reliability of coral records from the western Pacific warm pool: A comparison using age-optimized records. *Paleoceanography* **4**, 457–464 (1999).
- Picaut, J., Ioualalen, L., Menkes, C., Delcroix, T. & McPhaden, M. Mechanism of the zonal displacements of the Pacific warm pool: Implications for ENSO. *Science* **274**, 1486–1489 (1996).
- Trenberth, K. E. & Shea, D. J. On the evolution of the Southern Oscillation. *Mon. Weath. Rev.* **115**, 3078–3096 (1987).
- Wolter, K. & Timlin, M. S. in *Proc. 17th Climate Diagnostics Workshop* 52–57 (Univ. of Oklahoma, Norman, Oklahoma, 1993).
- Gagan, M. K., Chivas, A. R. & Isdale, P. J. High-resolution isotopic records from corals using ocean temperature and mass spawning chronometers. *Earth Planet. Sci. Lett.* **121**, 549–558 (1994).
- Woodruff, S. D., Slutz, R. J., Jenne, R. L. & Steurer, P. M. A comprehensive ocean-atmosphere data set. *Bull. Am. Meteorol. Soc.* **68**, 1239–1250 (1987).
- Mann, M. E. & Lees, J. Robust estimation of background noise and signal detection in climatic time series. *Clim. Change* **33**, 409–445 (1996).
- Hughen, K. A., Schrag, D. P. & Jacobsen, S. B. El Niño during the last interglacial period recorded by a fossil coral from Indonesia. *Geophys. Res. Lett.* **36**, 3129–3132 (1999).

- Mann, M. E., Bradley, R. S. & Hughes, M. K. in *El Niño and the Southern Oscillation: Multiscale Variability and its Impacts on Natural Ecosystems and Society* (ed. Diaz, H. & Markgraf, V.) 321–372 (Cambridge Univ. Press, Cambridge, in the press).
- Mann, M. E., Bradley, R. S. & Hughes, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
- Wellington, G. M., Dunbar, R. B. & Merlen, G. Calibration of stable oxygen isotope signatures in Galapagos corals. *Paleoceanography* **11**, 467–480 (1996).
- Ropelewski, C. F. & Jones, P. D. An extension of the Tahiti–Darwin Southern Oscillation Index. *Mon. Weath. Rev.* **115**, 2161–2165 (1987).

Acknowledgements

We thank B. Vaughn, T. Kimbell, J. Thurston and the Kiribati Ministry for Natural Resources Development (particularly J. Uan) for logistical assistance. Laboratory analyses were facilitated by C. Evans, J. Kelleher and B. Vaughn. We thank J. White for discussions. Reviews by M. Evans and R. Dunbar greatly improved the manuscript. COADS data are provided by the NOAA-CIRES Climate Diagnostics Centre from their web site at www.cdc.noaa.gov, and with assistance from S. Worley. TAO array data are provided by the TAO Project Office web site at www.pmel.noaa.gov/toga-tao. We thank M. Mann and E. Cook for providing spectral analysis software and advice. This study was supported by the Earth System History Program of the US National Science Foundation and the NOAA Paleoclimatology Program.

Correspondence and requests for materials should be addressed to J.C. (e-mail: jcole@geo.arizona.edu). The bimonthly Maiana $\delta^{18}\text{O}$ record is available at <http://www.ngdc.noaa.gov/paleo/pubs/urban2000>.

Widespread uplift and ‘trapdoor’ faulting on Galápagos volcanoes observed with radar interferometry

Falk Amelung^{*†}, Sigurjón Jónsson^{*}, Howard Zebker & Paul Segall

Stanford University, Department of Geophysics, Mitchell Building, Stanford, California 94305-2215, USA

^{*}These authors contributed equally to this work.

Volcanic uplift, caused by the accumulation of magma in subsurface reservoirs, is a common precursor to eruptions^{1,2}. But, for some volcanoes, uplift of metres or more has not yet led to an eruption³. Here we present displacement maps of volcanoes in the Galápagos Islands, constructed using satellite radar interferometry, that might help explain this dichotomy. We show that all but one of the seven volcanoes on the islands of Isabela and Fernandina deformed during 1992–99. Cerro Azul and Fernandina erupted^{4–6} during the observation period and show evidence of inflation, co-eruptive deflation and shallow dyke intrusion. In contrast, the largest volcano, Sierra Negra, has not erupted, yet exhibits spatially and temporally variable deformation, with a maximum uplift of 2.7 m between 1992 and 1999, which can be modelled by a shallow inflating sill. Inflation during 1997–98, however, was accompanied by ‘trapdoor’ faulting on a steeply dipping fracture system within the caldera. Repeated trapdoor faulting over geological time has formed an arcuate intra-caldera ridge within Sierra Negra and may have acted to relax stresses above the magma chamber, inhibiting summit eruptions. Similar processes may help explain large uplift unaccompanied by eruptive activity at other volcanoes.

A mosaic of two deformation interferograms spanning 1992–7 and 1992–8 shows that all volcanoes on the islands Fernandina and Isabela, except Ecuador, deformed during the observation period (Fig. 1). Ecuador is also the only volcano not to have erupted in historic times (see Table 1). Before these observations there were no instrumental measurements of active deformation in the Galápagos,

[†] Present address: Hawaii Institute of Geophysics and Planetology, SOEST, University of Hawaii, 2525 Correa Road, Honolulu, Hawaii 96822, USA.

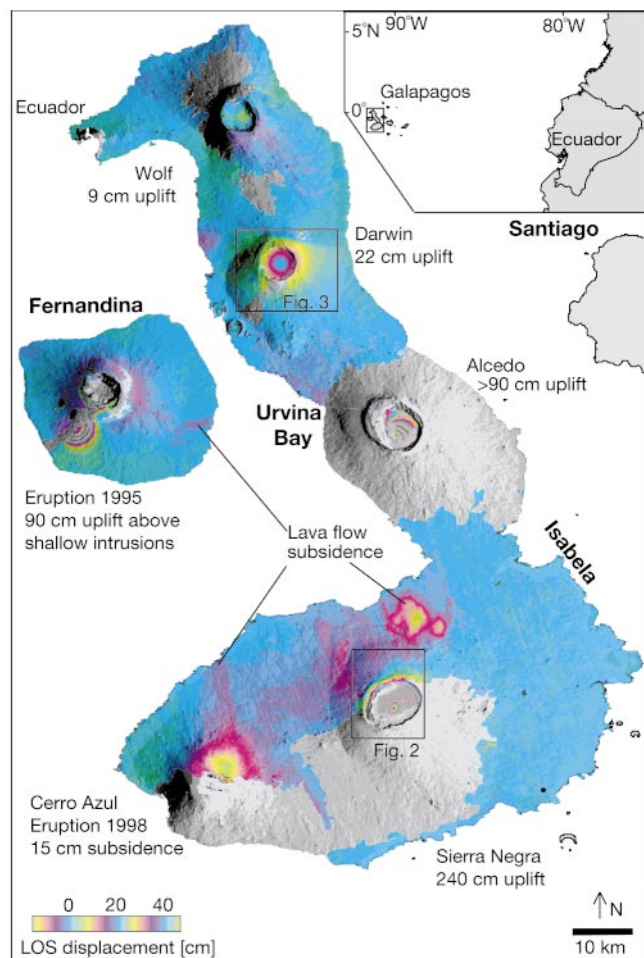


Figure 1 Observed deformation at Isabela and Fernandina (main image), the westernmost islands in the Galápagos archipelago (inset). The map shows line-of sight (LOS) ground displacements towards the European radar satellites ERS-1 and ERS-2, produced from interferograms spanning 5 years from 1992 to 1997 (Ecuador, Wolf, Fernandina) and 6.4 years from 1992 to 1998 (Darwin, Alcedo, Sierra Negra and Cerro Azul; Table 1). The radar echoes are reduced to interferograms by calculating the phase difference between images from two satellite passes^{22–24}. The topographic contribution to the observed phase was removed using a digital elevation model²⁵. The interferometric phase, which is measured modulo 2π , is unwrapped to²⁶ produce LOS displacement, and presented here so that each colour-cycle represents 20 cm displacement. The incidence angle of the ERS radars is 23° from vertical. Areas where the interferometric coherence is lost, presumably because of dense vegetation, are shown grey. Also shown is the maximum uplift at each volcano, assuming that displacements are vertical. All the volcanoes actively deformed during the observation period except Ecuador volcano. The 1995 Fernandina eruption occurred on the southwest flank. The 1998 Cerro Azul eruption occurred inside the summit caldera and on the vegetated southeast flank⁴.

although 5–7 m of uplift occurred in Urvina Bay in 1954, six months before an eruption of Sierra Negra^{7,8}. We derived uplift rates ranging from several millimetres per year at Wolf to 0.9 m per year at Sierra Negra using satellite radar interferometry (InSAR); these rates indicate accumulation of magma in reservoirs beneath the summit calderas of each volcano. Only Cerro Azul and Fernandina erupted during the observation period. At Cerro Azul, pre-eruptive uplift (June 1992 to October 1997), as well as subsidence (October 1997 to November 1998), are associated with the September–October 1998 eruption^{4,5}. Fernandina exhibits a semi-circular fringe pattern on its southwest flank (Fig. 1), which has been modelled with a dipping, radial dyke coincident with the 1995

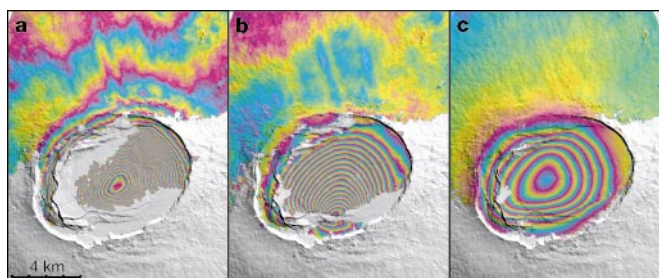


Figure 2 Radar interferograms of Sierra Negra volcano showing uplift during three time periods. **a**, 1992–97 (5.3 years, descending orbit). **b**, 1997–98 (1.1 years, descending orbit). **c**, 1998–99 (0.5 years, ascending orbit). Each colour cycle represents 5 cm LOS displacement. The deformation pattern during the first and third time period is similar, but is markedly different during the intervening period. In **b** the interferometric phase is discontinuous across the pre-existing fault (arcuate ridge) in the southern part of the caldera, the location of the proposed trapdoor faulting.

eruption⁶. Uplift of 0.1 m from September 1998 to March 1999 suggests refilling of a shallow magma chamber. Uplift at some volcanoes (for example, Campi Flegrei⁹ and Yellowstone¹⁰) may be related to hydrothermal rather than magmatic processes. Because of the high eruption frequency of the Galápagos volcanoes, and our observations of pre-eruptive uplift and co-eruptive subsidence at Cerro Azul, we suggest that the uplift at the various volcanoes in the Galápagos is due to magma accumulation.

The interferograms also show surprisingly large evidence for subsidence of lava flows on Fernandina, Cerro Azul and Sierra Negra (see Fig. 1). Up to 20 cm of subsidence was observed on the 1979 flows on the north flank of Sierra Negra (between 1992 and 1998), roughly 13–19 years after their emplacement¹¹. The subsidence on the north flank of Cerro Azul and on the southeast flank of Fernandina is found on lava flows of unknown age. Potential explanations for the subsidence include thermal contraction, and compaction of the flow, or lava substrate^{12–14}.

Sierra Negra experienced particularly dramatic and variable ground deformation. We observe a peak radar line-of-sight (LOS) displacement (range decrease) of 2.2 m between 1992 and 1998 (corresponding to 2.4 m of uplift; see Fig. 1) and 0.3 m during the September 1998 to March 1999 period. The corresponding displacement rates vary from 0.32 m yr^{-1} during 1992–97, to 0.90 m yr^{-1} during 1997–98, and 0.65 m yr^{-1} during 1998–99. Figure 2 illustrates that while the spatial pattern of deformation in 1992–97 and 1998–99 was similar, with the maximum uplift near the centre of the caldera, the deformation pattern during the intermediate period (1997–98) was markedly asymmetric with maximum uplift near the south rim of the caldera. Discontinuities in interferometric phase and loss of image coherence coincide with a previously mapped inner caldera fracture system (Fig. 2b). The fracture system, consisting of outward-dipping ($60\text{--}90^\circ$) normal faults, is marked by a 14-km-long curving arcuate ridge, open to the east, which surrounds a tilted and elevated inner caldera floor¹¹. We interpret the InSAR data as indicative of faulting along the southern part of the pre-existing fault system during the 1997–98 period. We observed similar fringe discontinuities near intra-caldera faults on Alcedo during 1992–98, but lack of data across the faults and on the volcano's flanks hindered a more detailed analysis.

We inverted the InSAR observations for magma chamber geometry and volume change in an elastic half-space using robust nonlinear optimization methods¹⁵, ignoring second-order effects of elastic and structural heterogeneity, and topography. For Darwin, Wolf and Cerro Azul, point centres of dilation¹⁶ (“Mogi sources”) accurately reproduce the near-radial symmetry observed in the fringe pattern (Fig. 3). These models are consistent with magma accumulation in shallow chambers 2–3 km beneath the calderas,

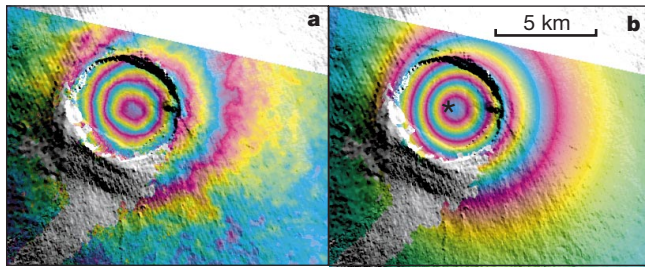


Figure 3 Deformation at Darwin volcano during 1992–98. **a**, Radar interferogram of Darwin volcano showing 0.2 m of LOS displacement (uplift). **b**, Predicted interferogram from the best-fit Mogi model with source location in the centre of the caldera (star) and depth of 3 km. Each colour cycle represents 5 cm LOS displacement. Observations at Wolf and Cerro Azul during 1992–97/98, and at Fernandina during 1998–99 are also well fitted with Mogi models at depths of approximately 2, 5 and 3 km, respectively.

except at Cerro Azul which yields a source depth of about 5 km. Shallow magma accumulation is compatible with inferred depths of fractional crystallization of evolved, well-mixed basalts^{17,18}. In contrast with Darwin, Wolf and Cerro Azul, simple point-source models do not accurately describe the data from Sierra Negra or Alcedo. The elliptical deformation pattern and absence of fringes outside the caldera at Sierra Negra require a more realistic magma chamber model than a Mogi source (Fig. 4a and b).

The radar data alone cannot uniquely resolve the shape of what are undoubtedly complex magma chambers, but we can fit the data with geologically plausible sources of deformation. The Sierra Negra data from the first and third time periods (Fig. 2a and c) can be fitted with a sill whose opening varies spatially. We first assume a horizontal sill with uniform opening and solve for the length, width, depth and opening. We retain the optimal depth of 1.9 km, enlarge the surface area, and solve for the spatial distribution of opening. We find a distribution of opening that fits the radar data in a least-squares sense, is reasonably smooth, and constrains the opening to be non-negative¹⁹. We found openings of up to 0.5 m between September 1998 and March 1999 (Fig. 4c and d). The magma body itself may be much thicker as the data are only sensitive to the change in thickness. This model is geologically reasonable and provides a much better fit to the observations than the point-source model (Fig. 4).

The 1997–98 observations at Sierra Negra (Fig. 2b) require a combination of two processes: uplift due to the expanding magma body and faulting along the inner caldera fracture system. We model the data with uniform sill opening and allow slip on a 75° dipping normal fault coincident with the mapped fracture system¹¹ (Fig. 5). We found fault slip of up to 1.2 m; the estimated seismic moment would correspond to an earthquake of magnitude 5.7 (ref. 20),

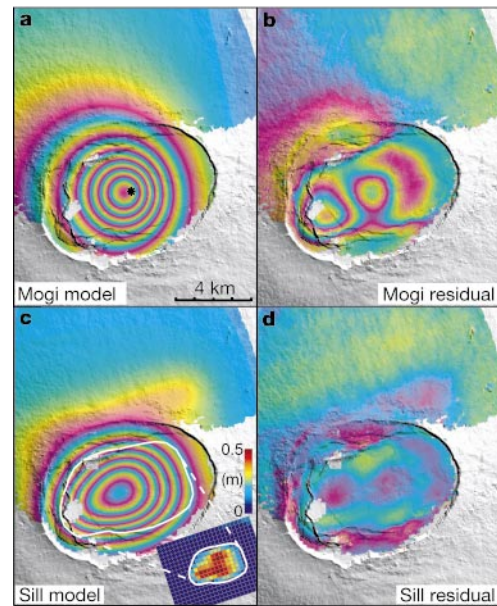


Figure 4 Two models for the Sierra Negra deformation during 1998–99. See data in Fig. 2c. **a**, Best-fit Mogi model. **b**, The residual, the difference between the data and the Mogi model prediction; large residual indicates a poor fit. **c**, Model of a horizontal sill at 1.9 km depth with variable openings of up to 0.5 m (shown in the lower right corner). **d**, The corresponding residual between the data and the sill model. The deformation on Sierra Negra from 1992–97 (Fig. 2a) is likewise better described by a sill than by a point-source model.

assuming a shear modulus of 30 GPa. The largest reported earthquake in the Sierra Negra area during this time interval was a $M_w = 5.0$ event on 11 January 1998 (see data from the National Earthquake Information Center (NEIC) at <http://www.neic.cr.usgs.gov/>). Unless the shear modulus of the near surface rocks is an order of magnitude less than assumed, this earthquake accounts for only 8% of the estimated moment release, suggesting that the slip was predominantly aseismic. Fault slip allowed the crust above the magma body to hinge upward like a trapdoor, with the sill opening occurring north of the normal fault system.

Uplift before 1997–98 caused extension of the caldera floor, eventually leading to trapdoor faulting sometime in 1997–98. The stress perturbations caused by magma chamber pressurization were thus sufficient to initiate slip on the fault, although conditions for a dyke to propagate to the surface were not achieved. Slip on the arcuate fault system apparently relaxed shear stresses along the fault because slip was not observed between 1998 and 1999, although uplift continued. Cumulative vertical offsets along the arcuate fault structure observed in the topography range from 10 m in the area of

Table 1 Maximum line-of-sight displacements at various volcanoes

Volcano	Last eruption	Track 140 6/15/92–10/16/97 5.3 years $B_{\perp} = 453$ m	Track 140 10/16/97–11/5/98 1.1 years $B_{\perp} = 367$ m	Track 140 6/15/92–11/5/98 6.4 years† $B_{\perp} = 85$ m	Track 412 9/12/92–9/30/97 5 years* $B_{\perp} = 279$ m	Track 61 9/26/98–3/20/99 0.5 years $B_{\perp} = 5$ m
Ecuador	Not known ²¹	–	–	–	0	–
Wolf	1982 ⁸	–	–	–	0.08 m	–
Fernandina	1995 ⁶	–	–	–	0.85 m†	0.08 m
Darwin	1812 ⁸	0.17 m	<0.05 m	0.20 m	0.14 m	<0.03 m
Alcedo	~1946–1960 ⁷	>0.80 m	<0.05 m	>0.85 m	–	<0.03 m
Sierra Negra	1979 ¹¹	1.59 m	0.90 m	2.20 m	–	0.30 m
Cerro Azul	1998 ⁵	0.13 m	–0.26 m†	–0.14 m*	–	<0.03 m

Data was observed in five different interferograms derived from nine available radar scenes acquired in 1992 and in 1997–99. $B_{\perp}$ is the perpendicular component of the spatial separation between two satellite passes. Images from track 140 and 412 are acquired from descending satellite orbits and the images from track 61 from ascending orbits.

* In Fig. 1.

† Eruption occurred during this time period.

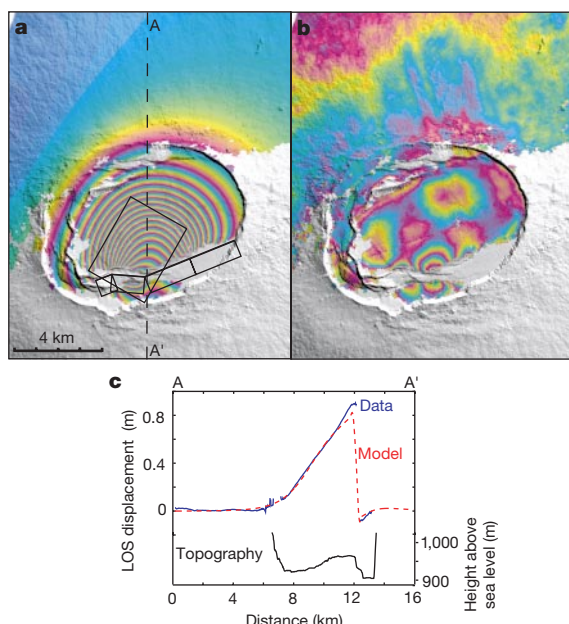


Figure 5 Model for the observed deformation on Sierra Negra during 1997–98. See Fig. 2b. **a**, The model has uniform opening of a horizontal sill (large rectangle) and slip on the normal faults that bound the elevated inner caldera floor. The four smaller rectangles show the surface projection of normal fault patches used in the modelling; a thick line marks the upper edge. The depth of the sill is not well constrained in this case; estimates range between 2.3 and 2.9 km. The difference between this and the depth obtained when trapdoor faulting was not present (2.2 km in 1992–97 and 1.9 km in 1998–99) is unlikely to be significant owing to imperfect resolution and model approximations. **b**, Residual between the data and the model prediction. Localized fringes near the fault are due to the simplified representation of the fault as four rectangular dislocation segments. **c**, Comparison between the data and the model showing the LOS displacement along the profile A–A' as well as the intra-caldera surface topography. The observed displacement follows the topography indicating repeated slip events.

maximum uplift in 1997–98 to over 100 m in the western caldera. In places the ridge actually extends above the caldera rim, indicating that trapdoor faulting has occurred numerous times since the caldera floor was resurfaced. The last four eruptions (and perhaps more) were not intra-caldera eruptions but occurred in the area north of the caldera. Our observations suggest that repeated slip on intra-caldera (or caldera bounding) faults may allow uplift without accumulating sufficient stress to permit vertical dyke propagation. Similar processes may explain large uplifts without eruptions at other rapidly deforming volcanoes³. □

Received 3 March; accepted 24 August 2000.

1. Dvorak, J. J. & Dzurlin, D. Volcano geodesy: the search for magma reservoirs and the formation of eruptive vents. *Rev. Geophys.* **35**, 343–384 (1997).
2. Lipman, P. W., Moore, J. G. & Swanson, D. A. Bulging of the north flank before the May 18 eruption—geodetic data. *USGS Prof. Pap.* **1250**, 143–155 (1981).
3. Newhall, C. G. & Dzurlin, D. Historical unrest at large calderas of the world. *USGS Bull.* **1855**, 509–520 (1988).
4. Naumann, T. & Geist, D. Physical volcanology and structural development of Cerro Azul Volcano, Isabela Island, Galápagos: Implication for the development of Galápagos-type shield volcanoes. *Bull. Volcanol.* **61**, 497–514 (2000).
5. Mouginiis-Mark, P. J., Snell, H. & Ellis, R. GOES satellite observations of the 1998 eruption of Volcan Cerro Azul, Galápagos Islands. *Bull. Volcanol.* **62**, 188–198 (2000).
6. Jónsson, S. *et al.* A shallow-dipping dike fed the 1995 flank eruption at Fernandina volcano, Galápagos, observed by satellite radar interferometry. *Geophys. Res. Lett.* **26**, 1077–1080 (1999).
7. Geist, D., Howard, K. A., Jellinek, A. M. & Rayder, S. The volcanic history of Volcán Alcedo, Galápagos Archipelago: A case study of rhyolitic oceanic volcanism. *Bull. Volcanol.* **56**, 243–260 (1994).
8. Simkin, T. & Siebert, L. *Volcanoes of the World* 2nd edn (Geoscience, Tucson, 1994).
9. Bonafede, M. & Mazzanti, M. Modelling gravity variations consistent with ground deformation in the Campi Flegrei caldera (Italy). *J. Volcanol. Geotherm. Res.* **81**, 137–157 (1998).
10. Wicks, C. Jr, Thatcher, W. & Dzurlin, D. Migration of fluids beneath Yellowstone caldera inferred from satellite radar interferometry. *Science* **282**, 458–462 (1998).

11. Reynolds, R. W., Geist, D. & Kurz, M. D. Physical volcanology and structural development of Sierra Negra volcano, Isabela island, Galápagos archipelago. *Geol. Soc. Am. Bull.* **107**, 1398–1410 (1995).
12. Murray, J. B. The influence of loading by lavas on the siting of volcanic eruption vents on Mt Etna. *J. Volcanol. Geotherm. Res.* **35**, 121–139 (1988).
13. Briole, P., Massonnet, D. & Delacourt, C. Post-eruptive deformation associated with the 1986–87 and 1989 lava flows on Etna detected by radar interferometry. *Geophys. Res. Lett.* **24**, 37–40 (1997).
14. Stevens, N. F. Surface movements of emplaced lava flows measured by SAR interferometry. *J. Geophys. Res.* (in the press).
15. Cervelli, P., Murray, M. H., Segall, P., Aoki, Y. & Kato, T. Estimating source parameters from deformation data, with an application to the March 1997 earthquake swarm off the Izu Peninsula, Japan. *J. Geophys. Res.* (in the press).
16. Mogi, K. Relations between the eruptions of various volcanoes and the deformation of the ground surface around them. *Bull. Earth. Res. Inst. Univ. Tokyo* **36**, 99–134 (1958).
17. Geist, D., Naumann, T. & Larson, P. Evolution of Galápagos magmas: Mantle and crustal level fractionation without assimilation. *J. Petrol.* **39**, 953–971 (1998).
18. Allen, J. F. & Simkin, T. Fernandina Volcano's evolved, well-mixed basalts: Mineralogical and petrological constraints on the nature of the Galápagos plume. *J. Geophys. Res.* **105**, 6017–6041 (2000).
19. Lawson, C. L. & Hanson, R. J. *Solving Least Squares Problems* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974).
20. Kanamori, H. & Anderson, D. L. Theoretical basis of some empirical relations in seismology. *Bull. Seismol. Soc. Am.* **65**, 1073–1095 (1975).
21. Rowland, S., Munro, D. C. & Perez-Oviedo, V. Volcán Ecuador, Galápagos islands: erosion as a possible mechanism for the generation of steep-sided basaltic volcanoes. *Bull. Volcanol.* **56**, 271–283 (1994).
22. Massonnet, D. *et al.* The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* **364**, 138–142 (1993).
23. Zebker, H., Rosen, P., Goldstein, R., Gabriel, A. & Werner, C. On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *J. Geophys. Res.* **99**, 19617–19643 (1994).
24. Massonnet, D. & Feigl, K. Radar interferometry and its application to changes in the Earth's surface. *Rev. Geophys.* **36**, 441–500 (1998).
25. Mouginiis-Mark, P. J., Rowland, S. K. & Garbeil, H. Slopes of western Galápagos volcanoes from airborne interferometric radar. *Geophys. Res. Lett.* **23**, 3767–3770 (1996).
26. Goldstein, R. M., Zebker, H. A. & Werner, C. L. Satellite radar interferometry: Two-dimensional unwrapping. *Radio Sci.* **23**, 713–720 (1988).

Acknowledgements

We thank the European Space Agency for providing the SAR data and for supporting F.A. with a post-doctoral fellowship. This study was also supported by NASA. We thank D. Geist and P. Lundgren for comments.

Correspondence and requests for materials should be addressed to S.J. (e-mail: jonsson@pangea.stanford.edu) or F.A. (e-mail: amelung@pgd.hawaii.edu).

Unicellular C₄ photosynthesis in a marine diatom

John R. Reinfelder*, Anne M. L. Kraepiel† & François M. M. Morel‡

* Department of Environmental Sciences, Rutgers University, 14 College Farm Road, New Brunswick, New Jersey 08901, USA

† Centre de Geochemie de la Surface, Université de Strasbourg, 1 rue Blessig, 67084 Strasbourg, France

‡ Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA

Nearly 50 years ago, inorganic carbon was shown to be fixed in microalgae as the C₃ compound phosphoglyceric acid¹. The enzyme responsible for C₃ carbon fixation, ribulose-1,5-bisphosphate carboxylase (Rubisco), however, requires inorganic carbon in the form of CO₂ (ref. 2), and Rubisco enzymes from diatoms have half-saturation constants for CO₂ of 30–60 μM (ref. 3). As a result, diatoms growing in seawater that contains about 10 μM CO₂ may be CO₂ limited⁴. Kinetic and growth studies have shown that diatoms can avoid CO₂ limitation^{5–7}, but the biochemistry of the underlying mechanisms remains unknown. Here we present evidence that C₄ photosynthesis supports carbon assimilation in

the marine diatom *Thalassiosira weissflogii*, thus providing a biochemical explanation for CO₂-insensitive photosynthesis in marine diatoms. If C₄ photosynthesis is common among marine diatoms, it may account for a significant portion of carbon fixation and export in the ocean, and would explain the greater enrichment of ¹³C in diatoms compared with other classes of phytoplankton. Unicellular C₄ carbon assimilation may have predated the appearance of multicellular C₄ plants.

There is a long-standing debate regarding the existence of a C₄ photosynthetic pathway in marine phytoplankton^{8–10}. In 'classic' multicellular C₄ plants, carbon acquisition is facilitated by the enzyme phosphoenolpyruvate carboxylase (PEPCase), which catalyses the carboxylation of PEP (using HCO₃[–] as the inorganic carbon substrate^{11,12} instead of CO₂) leading to the synthesis of a C₄ compound—malic or aspartic acid. Enzymatic decarboxylation of

the C₄ compound by malic enzyme or phosphoenolpyruvate carboxykinase (PEPCKase) then generates CO₂ for fixation by Rubisco in the first step of the Calvin cycle¹³. PEPCase and PEPCKase activities have been measured in marine diatoms^{8,10}

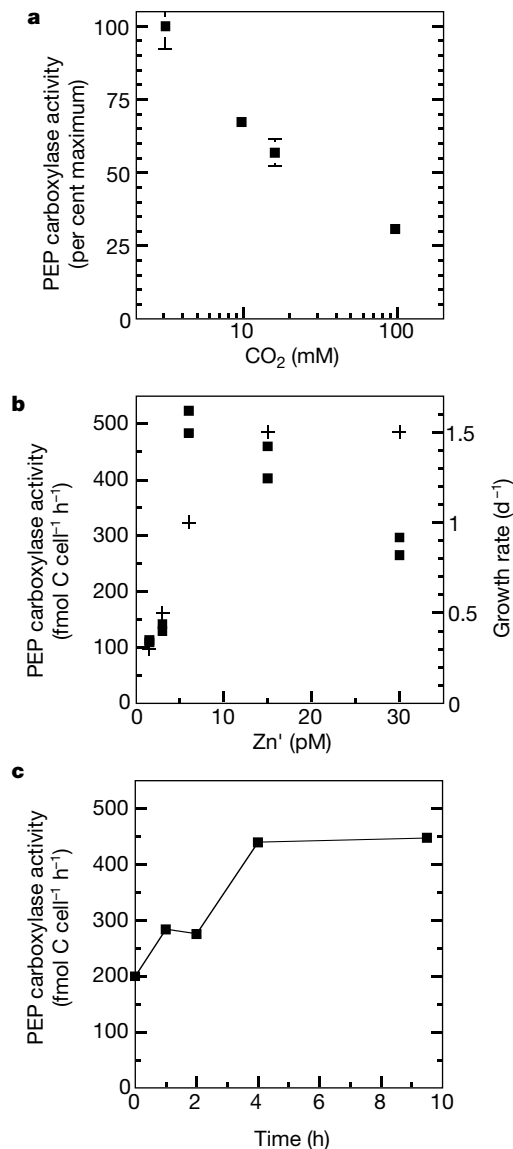


Figure 1 Phosphoenolpyruvate carboxylase (PEPCase) activities (squares) and growth rates (+) in the marine diatom *T. weissflogii*. **a**, *T. weissflogii* acclimated to a range of CO₂ concentrations. PEPCase activities are the means of triplicate analyses. **b**, *T. weissflogii* acclimated to a range of inorganic Zn (Zn²⁺) concentrations. PEPCase activities are the range of activities in replicate cultures. **c**, PEPCase activities in Zn-replete cells exposed to the carbonic anhydrase (CA) inhibitor ethoxzolamide (100 μM). These cells grew at a rate of 1 d^{–1} and had a photosynthetic rate (measured by short-term ¹⁴C fixation) of 520 fmol C cell^{–1} h^{–1}.

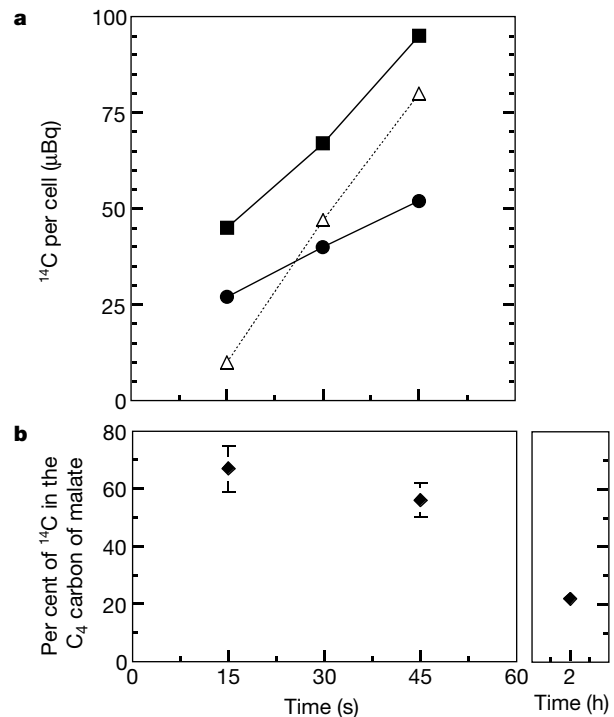


Figure 2 Short-term ¹⁴C assimilation in the marine diatom *T. weissflogii*. **a**, Radiocarbon contents (1 μBq = 10^{–6} disintegrations per second) of malate (squares), phosphoglyceric acid (circles) and sugars (triangles) in cells exposed to ¹⁴C for 45 s in pH-buffered medium containing 0.4 mM total inorganic carbon. Error bars for 2σ propagated counting errors are smaller than plot symbols. **b**, Per cent of ¹⁴C in the C₄ carbon of malate in diatom cells during short-term labelling and after 2 h exposure to ¹⁴C with a total inorganic carbon concentration of 1.8 mM. 2σ propagated counting errors were ≥ the difference between replicate analyses.

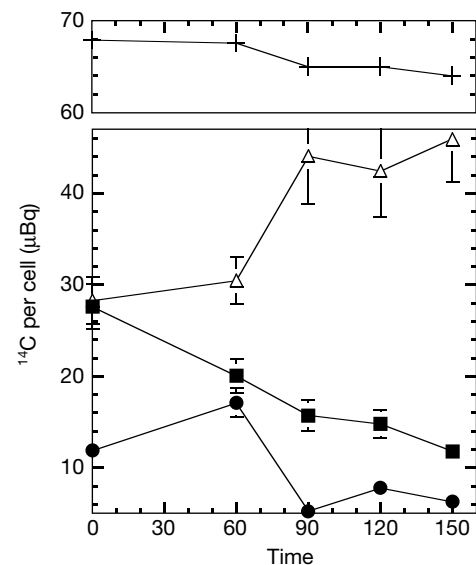


Figure 3 Transfer of ¹⁴C from malate to C₃ compounds (PGA and sugars) in the marine diatom *T. weissflogii*. Data are the radiocarbon contents (1 μBq = 10^{–6} disintegrations per second) of malate (squares), phosphoglyceric acid (circles), sugars (triangles) and total acid-stable ¹⁴C (+) in cells exposed to ¹⁴C for 60 s in pH-buffered medium containing 1.8 mM total inorganic carbon and then transferred to ¹²C medium containing 2 mM inorganic carbon.

but, on the basis of short-term ^{14}C labelling studies that showed a lack of significant synthesis of the C_4 compounds malate or aspartate, it was concluded that PEP carboxylation may be involved in the synthesis of amino-acid precursors but it does not lead to carbon fixation through the Calvin cycle^{8,9}.

Previous studies might have missed the importance of PEPCase in diatom photosynthesis partly because culture conditions placed no burden on the carbon acquisition systems of the cells. We found that PEPCase activity in a coastal Atlantic Ocean clone (ACTIN) of the cosmopolitan marine diatom *T. weissflogii* increases in cells acclimated to low CO_2 (Fig. 1a), indicating a probable link between inorganic carbon uptake and PEPCase. If this is the case, we would also expect an increase in PEPCase activity in Zn-stressed cells. At low Zn, a decreased concentration and activity of the Zn enzyme carbonic anhydrase (CA), which reversibly catalyses the interconversion of CO_2 and HCO_3^- impairs inorganic carbon accumulation in *T. weissflogii*^{14,15}. We therefore examined PEPCase activity in *T. weissflogii* acclimated to a range of inorganic Zn concentrations (Zn') that are known to impose stress on the carbon acquisition system. The highest PEPCase activities are measured in *T. weissflogii* cells cultured with 6 pM Zn' (Fig. 1b), whose growth rate (1 d^{-1}) is slightly reduced relative to Zn-sufficient cells (1.5 d^{-1}). At very low Zn' , PEPCase activity declines in proportion to the growth rate, which becomes Zn limited. Perhaps fortuitously, the measured PEPCase activity in low Zn cultures ($110\text{ fmol cell}^{-1}\text{ h}^{-1}$ at $\text{Zn}' = 1.5\text{ pM}$) is about 90% of the carbon fixation rate in cells growing at a rate of 0.3 d^{-1} ($0.3\text{ d}^{-1} \times 10\text{ pmol cell}^{-1}/24\text{ h d}^{-1} = 125\text{ fmol cell}^{-1}\text{ h}^{-1}$). In cells grown with 6, 15 and 30 pM Zn' , measured PEPCase activity is about 120%, 70% and 45% of carbon fixation, respectively. Increased PEPCase activity in cells subjected to moderate Zn stress ($\text{Zn}' = 6\text{ pM}$) is probably a response to diminished intracellular and extracellular CA activity. Confirmation of this is seen in a similar increase in PEPCase activity measured in cells exposed to the CA inhibitor ethoxymalate (Fig. 1c).

If PEPCase supports carbon fixation by Rubisco in cells whose carbon acquisition system is compromised by either low CO_2 or Zn' , we would expect to find significant short-term fixation of carbon as the C_4 product of PEPCase in CO_2 or Zn-stressed *T. weissflogii*. In short-term photosynthesis experiments, after 10–15 s, the C_4 compound malate was the principal ^{14}C -labelled intracellular pool in low Zn ($\text{Zn}' = 4\text{ pM}$) diatoms (Fig. 2a) and almost twice as large as the labelled C_3 pool (phosphoglyceric acid; PGA) (malate:PGA = 1.8 ± 0.09 , $n = 3$). In diatoms grown with

higher Zn ($\text{Zn}' = 15\text{ pM}$), the short-term ^{14}C -labelled malate:PGA ratio was ≤ 1 and decreased from 1.0 ± 0.08 to 0.73 ± 0.06 as the concentration of aqueous CO_2 in the cultures was increased from 3 to 16 μM . All these results are consistent with major primary carbon fixation by PEPCase in cells that are stressed by either low concentrations of CO_2 ($< 3\text{ }\mu\text{M}$) or low Zn ($\text{Zn}' < 15\text{ pM}$). The observed β -carboxylation should lead to the short-term labelling of carbon in the C_4 position of malate. Indeed, *in vitro* decarboxylation of pulse-labelled malate from low Zn cells showed that nearly 70% of the ^{14}C accumulated in 15 s in malate was in the C_4 carbon (Fig. 2b). After 2 h this proportion decreased to about 25%, as expected for uniform label incorporation in malate (Fig. 2b).

We followed the fate of malate carbon in ^{14}C -labelled cells transferred to ^{12}C medium. During the first 60 s of this 'chase', there was a simultaneous decline in the ^{14}C activity of malate and an increase in the ^{14}C content of PGA and sugars (Fig. 3). Mass balance of ^{14}C during this period requires the transfer of carbon from malate to PGA; ^{14}C content of malate decreased by $8\text{ }\mu\text{Bq}$ while that of PGA increased by $5\text{ }\mu\text{Bq}$ and that of sugars increased by $2\text{ }\mu\text{Bq}$. Thus malate is being decarboxylated and the released CO_2 is being fixed by Rubisco to produce PGA and sugars. We have observed this pattern of ^{14}C transfer from malate to PGA in four experiments with different batches of cells (Table 1).

The operation of C_4 photosynthesis requires a decarboxylase in close proximity to Rubisco to insure efficient transfer of carbon. As PEPCKase activity has been measured in marine diatoms³, we reasoned that malate is decarboxylated by PEPCKase (after oxidation to oxaloacetate) in the chloroplast of *T. weissflogii*. We found 75% of PEPCKase activity in *T. weissflogii* located in chloroplasts (Table 2). As this is the same percentage as measured for the chloroplastic enzyme Rubisco (which is less than 100% because of chloroplast breakage during preparation), all the PEPCKase is probably in the chloroplast. In very low Zn cells ($\text{Zn}' = 1.5\text{ pM}$), we measured a PEPCKase activity ($170 \pm 10\text{ fmol C cell}^{-1}\text{ h}^{-1}$) that is of the order of carbon fixation rates ($125\text{ fmol C cell}^{-1}\text{ h}^{-1}$) expected at that Zn-limited growth rate.

In the cytoplasm of the mesophyll cells of multicellular C_4 plants, CA provides PEPCase with HCO_3^- through the hydration of CO_2 and the cytoplasmic localization of CA is characteristic of such cells¹⁶. In *T. weissflogii*, all of the observed intracellular CA activity is also in the cytoplasm, the remainder being extracellular CA (Table 2; A. M. L. K. *et al.*, manuscript in preparation). This suggests that CA provides HCO_3^- to PEPCase which should also be located in the

Table 1 Transfer of ^{14}C from malate to C_3 compounds (PGA and sugars) in the marine diatom *T. weissflogii*

	$\text{Zn}' = 4\text{ pM}$		$\text{Zn}' = 4\text{ pM}$		$\text{Zn}' = 15\text{ pM}$		$\text{Zn}' = 15\text{ pM}$	
	Pulse	Chase	Pulse	Chase	Pulse	Chase	Pulse	Chase
Duration	10 s	25 s	60 s	60 s	5 s	12 s	10 s	12 s
Malate	60	39	41	30	43	20	40	19
PGA	31	43	17	25	51	72	52	73
Sugars	9	18	42	45	6	8	8	8

Cells acclimated to either 4 or 15 pM inorganic Zn (Zn') were labelled with ^{14}C for 5–60 s (pulse) and then transferred to ^{12}C medium for 12–60 s (chase). Values are percent total organic ^{14}C present as malate, phosphoglyceric acid (PGA) or sugars.

Table 2 Distribution of enzyme activities among cellular fractions of the marine diatom *T. weissflogii*

Fraction	PEPCase	PEPCKase	Rubisco	CA ($\times 10^{-5}$)
Whole cell	78	97	175	1.49
Chloroplasts	41 (53)*	73 (75)*	136 (78)*	~0 (0)†
Cytoplasm	37 (47)†	24 (25)†	39 (22)†	0.81 (54)
Extracellular	n.d.	n.d.	n.d.	0.67 (45)

T. weissflogii were grown with 3 pM Zn' (PEPCase, PEPCKase, Rubisco) or 15 pM Zn' (CA). Values are activities in fmol C per cell per hour for PEPCase, PEPCKase and Rubisco, and enzyme units per cell as defined²⁶ for CA. Percentages of whole-cell activities are given in parentheses. n.d., not determined.

* Calculated as the product of measured chloroplast activity per mg Chla and mg Chla per whole cell.

† Calculated as the difference between measured whole-cell and chloroplast activities.

‡ Calculated as the difference between measured whole-cell activity and cytoplasmic plus extracellular activities.

cytoplasm to prevent futile competition with Rubisco. We found nearly 50% of PEPCase in the cytoplasm of *T. weissflogii* (Table 2). The PEPCase activity we measured in the chloroplast fraction of *T. weissflogii* is probably cytoplasmic enzyme that partitioned with the denser material of the chloroplasts¹⁷.

Although the exact mechanisms of inorganic carbon uptake and assimilation in diatoms, particularly the relative importance of the C₄ and C₃ pathways under various conditions, need to be further elucidated, our results show that C₄ photosynthesis accounts for a major proportion of carbon fixation in CO₂ or Zn-stressed cells of *T. weissflogii*. As in the aquatic monocot *Hydrilla verticillata*¹⁸, carboxylation and decarboxylation occur simultaneously in a single cell, but in separate intracellular compartments. Because marine diatoms are especially important in the biological removal of CO₂ from surface waters to the deep sea¹⁹, a significant portion of export production in the ocean may be fixed through a C₄ assimilation pathway. The use of C₄ photosynthesis may explain the enrichment of ¹³C in diatoms relative to other classes of marine phytoplankton²⁰, and the increase in the ¹³C content of sediment organic matter from 35 Myr (early Oligocene) to the present²¹. Evolutionarily, the appearance of C₄ photosynthesis in diatoms is probably an adaptation to carbon acquisition stress caused by low CO₂ and/or Zn concentrations in surface waters. Diatoms radiated during the Mesozoic²²—an era of low concentrations of atmospheric CO₂ relative to earlier eras (Precambrian, Paleozoic)²³ when most photosynthetic microorganisms evolved. In this case, C₄ photosynthesis in aquatic microalgae would be much more ancient than in terrestrial C₄ and Crassulacean acid metabolism plants where it arose only about 7 Myr ago (late Miocene)²⁴. □

Methods

Enzyme activities

We grew cultures of *T. weissflogii* (clone ACTIN) axenically in trace-metal-buffered, artificial seawater (Aquil²⁵) at each concentration of CO₂ or Zn' for at least two transfers (9–10 generations) and collected them during mid-exponential growth. Phosphoenolpyruvate carboxylase activities were measured in cells homogenized by sonication in 50 mM bicine buffer (0 °C, pH 7.5) containing 10 mM NaHCO₃, 1.5 M glycerol, 1 mM EDTA, 10 mM MgCl₂, 5 mM dithiothreitol (DTT) and 5 mg ml⁻¹ bovine serum albumin (BSA) by PEP-dependent ¹⁴C-fixation at 25 °C. Phosphoenolpyruvate carboxykinase activities were determined by PEP-dependent ¹⁴C fixation in the presence of ADP and Mn in cells homogenized in 25 mM HEPES buffer (pH 7.1), containing 10 mM NaHCO₃, 1.5 M glycerol, 0.2 mM EDTA, 2 mM MnCl₂, 2 mM ADP, 0.1 M KCl, 5 mM DTT and 5 mg ml⁻¹ BSA. Carbonic anhydrase was measured in homogenized whole cells (total activity), intact whole cells, (extracellular activity) and in cell-fractionation (see below) supernatants (cytoplasmic activity) by a pH drift assay²⁶.

Short-term ¹⁴C assimilation

Mid-exponential phase diatoms were concentrated 50–100 times, resuspended in 1 or 2 ml buffer (350 mM sorbitol, 10 mM bicine, pH 8) and incubated in a stirred cell in the light (400 μmol photons m⁻² s⁻¹). For short-term uptake studies cells were incubated in the light in a sealed oxygen electrode cell (Fig. 2a) until no net O₂ production was observed (10 min) indicating that all inorganic carbon had been depleted from the buffer, or in previously prepared low dissolved inorganic carbon buffer (Table 1, Zn' = 4 pM, 10 s pulse and Zn' = 15 pM). We began experiments by adding 5 μCi (0.5 ml) of ¹⁴C as aqueous NaH¹⁴CO₃ resulting in a total inorganic carbon concentration of 0.4 mM. For the malate labelling (Fig. 2b) and ¹⁴C chase (Fig. 3) experiments, cells were incubated in the light in an unsealed vial for about 2 min before the addition of 5 μCi of ¹⁴C, resulting in a total inorganic carbon concentration of 1.8 mM. For chase experiments, ¹⁴C-labelled cells were collected on a 3-μm polycarbonate filter (Fig. 3) or centrifuged through a silicone oil layer (Table 1, Zn' = 4 pM, 10 s pulse and Zn' = 15 pM), resuspended into buffer containing 2 mM total inorganic ¹²C (no ¹⁴C), and returned to the light. Cell samples (0.25 ml) were transferred to 2 ml CHCl₃:MeOH (2:1) for kill and extraction. The upper phase of each extract (H₂O:MeOH mixture) was diluted 1:1.25 with 0.3 M KH₂PO₄ (pH 2.8) and the MeOH evaporated. Radiolabelled organic compounds were separated by anion exchange chromatography using high performance liquid chromatography²⁷ and ¹⁴C was quantified by liquid scintillation counting.

Malate decarboxylation

We determined the radiocarbon content of the C₄ carbon of malate by enzymatic digestion²⁸ of malate collected in the short-term labelling experiment described above. Radiolabelled malate was incubated at 30 °C with 0.25 units of malic enzyme and 20 units of glutathione reductase (Sigma) in 0.1 M HEPES (pH 7.5) containing 40 μM NADP, 20 mM oxidized glutathione and 4 mM MnCl₂. Decarboxylation was carried out for 48 h

to ensure completeness at which time the reaction was stopped by the addition of 80 μl 6 M HCl. Samples were evaporated and remaining ¹⁴C quantified by liquid scintillation counting. The percentage of ¹⁴C in the C₄ carbon was calculated as 100 × [1 – (malate ¹⁴C post-decarboxylation / malate ¹⁴C pre-decarboxylation)].

Cell fractionation

Diatom chloroplasts were isolated from *T. weissflogii* cells homogenized with a Polytron blender (3 × 30 s, 15,000 r.p.m.) in 50 mM HEPES buffer (0 °C, pH 8) containing 626 mM sorbitol, 1 mM EDTA and 10 mg ml⁻¹ bovine serum albumin. Lysed cells were overlain on top of a 70% Percoll solution (626 mM sorbitol, 1 mg ml⁻¹ bovine serum albumin) and centrifuged for 4 min at 4,000 r.p.m. and 4 °C. Chloroplasts were collected from above the Percoll cushion. Cytoplasm for CA analysis was collected as the supernatant of the Percoll centrifugation.

Received 25 April; accepted 26 June 2000.

- Calvin, M. *et al.* Carbon dioxide assimilation in plants. *Symp. Soc. Exp. Biol.* **5**, 284–305 (1951).
- Cooper, T. G., Filmer, D., Wishnick, M. & Lane, M. D. The active species of "CO₂" utilized by ribulose diphosphate carboxylase. *J. Biol. Chem.* **244**, 1081–1083 (1969).
- Badger, M. R. *et al.* The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO₂-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052–1071 (1998).
- Riebesell, U., Wolf-Gladrow, D. A. & Smetacek, V. S. Carbon dioxide limitation of marine phytoplankton growth rates. *Nature* **361**, 249–251 (1993).
- Colman, B. & Rotatore, C. Photosynthetic inorganic carbon uptake and accumulation in two marine diatoms. *Plant Cell Environ.* **18**, 919–924 (1995).
- Korb, R. E., Saville, P. J., Johnston, A. M. & Raven, J. A. Sources of inorganic carbon for photosynthesis by three species of marine diatom. *J. Phycol.* **33**, 433–440 (1997).
- Tortell, P. D., Reinfelder, J. R. & Morel, F. M. M. Active uptake of bicarbonate by diatoms. *Nature* **390**, 243–244 (1997).
- Morris, I. in *Primary Productivity in the Sea* (ed. Falkowski, P. G.) 139–159 (Plenum, New York, 1980).
- Mortain-Bertrand, A., Descolas-Gros, C. & Jupin, H. Short-term ¹⁴C incorporation in *Skeletonema costatum* (Greville) Cleve (Bacillariophyceae) as a function of light regime. *Phycologia* **26**, 262–269 (1987).
- Falkowski, P. G. & Raven, J. A. *Aquatic Photosynthesis* (Blackwell, Malden, 1997).
- Cooper, T. G. & Wood, H. G. The carboxylation of phosphoenolpyruvate and pyruvate. *J. Biol. Chem.* **246**, 5488–5490 (1971).
- Ausenhuis, S. L. & O'Leary, M. H. Hydrolysis of phosphoenolpyruvate catalysed by phosphoenolpyruvate carboxylase from *Zea mays*. *Biochem.* **31**, 6427–6431 (1992).
- Hatch, M. D. C₄ photosynthesis: a unique blend of modified biochemistry, anatomy and ultra-structure. *Biochim. Biophys. Acta* **895**, 81–106 (1987).
- Morel, F. M. M. *et al.* Zinc and carbon co-limitation of marine phytoplankton. *Nature* **369**, 740–742 (1994).
- Lane, T. W. & Morel, F. M. M. Regulation of carbonic anhydrase expression by Zn, Co, and CO₂ in the marine diatom *Thalassiosira weissflogii*. *Plant Physiol.* **123**, 345–352 (2000).
- Hatch, M. D. & Burnell, J. N. Carbonic anhydrase activity in leaves and its role in the first step of C₄ photosynthesis. *Plant Physiol.* **93**, 825–828 (1990).
- Slack, C. R., Hatch, M. D. & Goodchild, D. J. Distribution of enzymes in mesophyll and parenchyma-sheath chloroplasts of maize leaves in relation to the C₄-dicarboxylic acid pathway of photosynthesis. *Biochem. J.* **114**, 489–498 (1969).
- Magnin, N. C., Cooley, B. A., Reiskind, J. B. & Bowes, G. Regulation and localization of key enzymes during the induction of Kranz-less, C₄-type photosynthesis in *Hydrilla verticillata*. *Plant Physiol.* **115**, 1681–1689 (1997).
- Boyd, P. & Newton, P. Evidence of the potential influence of planktonic community structure on the interannual variability of particulate organic carbon flux. *Deep-Sea Res.* **42**, 619–639 (1995).
- Fry, B. & Wainright, S. C. Diatom sources of ¹³C-rich carbon in marine food webs. *Mar. Ecol. Prog. Ser.* **76**, 149–157 (1991).
- Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of C-13 in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).
- Medlin, L. K., Kooistra, W. H. C. F., Gersonde, R., Sims, P. A. & Wellbrock, U. Is the origin of the diatoms related to the end-permian mass extinction? *Nova Hedwigia*. **65**, 1–11 (1997).
- Berner, R. A. Atmospheric carbon dioxide levels over phanerozoic time. *Science* **249**, 1382–1386 (1990).
- Ehleringer, J. R. & Monson, R. K. Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu. Rev. Ecol. Syst.* **24**, 411–439 (1993).
- Price, N. M. *et al.* Preparation and chemistry of the artificial algal culture medium Aquil. *Biol. Oceanogr.* **6**, 443–461 (1988/89).
- Wilbur, K. M. & Anderson, N. G. Electrometric and colourimetric determination of carbonic anhydrase. *J. Biol. Chem.* **176**, 147–154 (1948).
- Stream, M., McNeill, Y. R. & Yellowlees, D. Photosynthetic carbon dioxide fixation in zooxanthellae. *Mar. Biol.* **115**, 195–198 (1993).
- O'Leary, M. H., Rife, J. E. & Slater, J. D. Kinetic and isotopic effect studies of maize phosphoenolpyruvate carboxylase. *Biochem.* **20**, 7308–7314 (1981).

Acknowledgements

We wish to thank M. Spring for help with the enzyme analyses; S.I. Chang, C.-W. Fan; and P.D. Tortell for help with the pulse-chase experiments; I. Schaperdorth for help with the cell fractionation; and T. Lane and K. Keller for useful comments. Funding from the Center for Environmental Bioinorganic Chemistry supported by NSF and DOE.

Correspondence and requests for materials should be addressed to J.R.R. (e-mail: reinfelder@envsci.rutgers.edu).

Disruptive sexual selection for plumage coloration in a passerine bird

Erick Greene*, Bruce E. Lyon†, Vincent R. Muehler‡, Laurene Ratcliffe§, Steven J. Oliver|| & Peter T. Boag§

* Division of Biological Sciences, The University of Montana, Missoula, Montana 59812, USA

† Biology Department, University of California, Santa Cruz, California 95064, USA

‡ 428 West 46th Street 4B, New York, New York 10036, USA

§ Biology Department, Queen's University, Kingston, Ontario K7L 3N6, Canada

|| Boston University Marine Program, Woods Hole, Massachusetts 02543, USA

The theory of sexual selection was developed to explain the evolution of highly exaggerated sexual ornaments¹. Now supported by vast empirical evidence², sexual selection is generally considered to favour individuals with the most extreme trait expression^{2–4}. Here we describe disruptive selection on a sexual ornament, plumage coloration, in yearling male lazuli buntings (*Passerina amoena*). In habitats with limited good-quality nesting cover, the dullest and the brightest yearlings were more successful in obtaining high-quality territories, pairing with females and siring offspring, than yearlings with intermediate plumage. This pattern reflects the way that territorial adult males vary levels of aggression to influence the structure of their social neighbourhood. Adult males showed less aggression towards dull yearlings than intermediate and bright ones, permitting the dull yearlings to settle on good territories nearby. Fitness comparisons based on paternity analyses showed that both the adults and dull yearlings benefited genetically from this arrangement, revealing a rare example of sexually selected male–male cooperation^{5,6}.

Lazuli buntings are socially monogamous songbirds that breed in early successional habitats throughout western North America⁷. At our study sites near Missoula, Montana (46° 48' N 114° 57' W), male buntings show dramatic variation in plumage coloration and territory quality, two characteristics that females may seek in their prospective mates^{2,8}. We studied the ecological and social consequences of this variation from 1992–1997. Males defend territories that include patches of shrubs in which the nests are built. Nesting cover is extremely patchy, ranging from single isolated shrubs to extensive areas of dense bushes. As a result, territory characteristics varied greatly among males: percentage shrub cover on territories ranged from 3 to 92%. Older males tend to return to the breeding grounds in the spring before yearling males, and competition among males for the limited areas with dense nesting cover can be intense^{7,9}.

Plumage coloration of males also varied dramatically, ranging from mainly dull brown, female-like plumage, to very bright blue, reddish, and white plumage (Fig. 1). We quantified this variation with a scoring system that yielded an overall index of each male's plumage brightness that is a composite of several measures (see Methods). Plumage scores ranged from 14 for extremely dull males to 36 for extremely bright males (see Fig. 1 for examples). Lazuli buntings exhibit "delayed plumage maturation" in which more than one year is generally required to attain full adult coloration^{10,11}. Thus, on average, yearling males were significantly duller than males of two or more years (hereafter called adults) (Fig. 1e). However, yearlings were also more variable than adults, so that some yearlings were extremely dull (Fig. 1a), whereas others (Fig. 1c) were almost as bright as adult males (Fig. 1d).

Plumage coloration had important reproductive consequences for yearling males (Fig. 2b, d, f), but not for adults (Fig. 2a, c, e). The

pairing success of yearlings showed a clear bimodal relation with plumage brightness: the dullest and brightest males obtained social partners while intermediate males had extremely low pairing success (Fig. 2b). Fitness analysis with univariate cubic splines¹² revealed significant disruptive sexual selection for yearling plumage coloration ($P < 0.01$). Disruptive selection is a rare and poorly understood mode of selection^{4,13} and has only been shown to occur twice for a sexually selected trait (body size¹⁴, pheromones¹⁵). Thus, identifying its causes is of broad interest.

Two factors combined to produce the disruptive selection for yearling plumage coloration: (1) the effect of plumage coloration on territory acquisition; and (2) strong female preference for high-quality territories, irrespective of a territorial male's plumage coloration. In yearling males, territory quality (estimated by per-

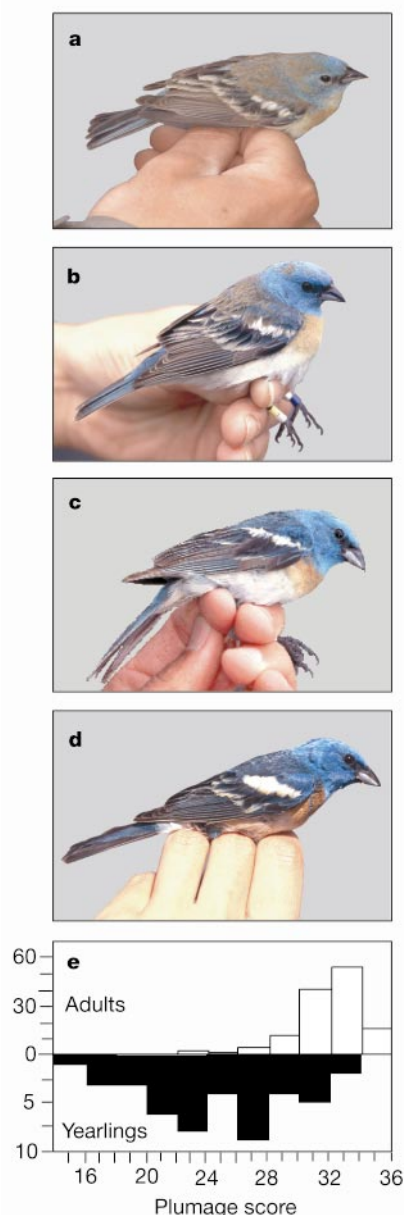


Figure 1 Variation in plumage coloration of male lazuli buntings. **a**, Dull yearling (plumage score, 17.5). **b**, Intermediate yearling (plumage score, 26). **c**, Bright yearling (plumage score, 31). **d**, Bright adult (plumage score, 34). **e**, Distribution of plumage scores of adult males (upper histogram) and yearling males (lower histogram). Yearling males have significantly lower plumage scores (24.9 ± 0.7 (mean $\pm$ s.e.m.), $n = 45$) than adults (32.1 ± 0.2 , $n = 129$; $t = 10.5$, $P < 0.001$; t -test); yearlings also have more variable plumage scores (variance ratio $F = 4.19$, $P < 0.01$).

centage nesting shrub cover) showed a disruptive pattern with plumage coloration (Fig. 2d) that closely parallels the bimodal pairing pattern (Fig. 2b): the duller and brightest yearling males generally obtained high-quality territories while intermediate yearlings settled in sparsely vegetated areas. This led to the bimodal pattern in pairing success because females preferentially paired with males on high-quality territories: paired males (all ages combined) had significantly more shrub cover on their territories ($47.0\% \pm 1.7$ (value $\pm$ standard error of the mean, s.e.m.), $n = 119$) than unpaired males (25.8 ± 3.4 , $n = 34$; Mann–Whitney z approximation is 4.97, $P < 0.0001$). These findings imply that territorial quality is an essential component of the disruptive selection, but raise the question of why bright and dull yearlings obtained good territories, while intermediate males did not.

The plumage coloration of yearling males influenced their ability to acquire territories by affecting their behavioural interactions with the more abundant adult males. During territory settlement, adult males were more aggressive to colourful males than to dull males (Fig. 3 and ref. 9). To examine the relation between the intensity of adult male aggression and the disruptive pairing and territory quality patterns that we observed for yearling males, we partitioned yearling plumage scores into three plumage classes on the basis of the disruptive pairing pattern (Fig. 2b). Dull plumage scored less than 24; intermediate plumage scored greater than or equal to 24 and less than 28; and bright plumage scored greater than or equal to 28 (other analyses below are also based on these categories). Adult male territory holders behaved significantly more aggressively towards bright and intermediate males than to dull yearling males (analysis of variance (ANOVA), $F_{3,55} = 8.95$, $P < 0.0001$). *Post hoc* paired comparisons (Fisher protected least significant difference (PLSD) tests) revealed significant differences between dull males and all other categories of males (all $P < 0.05$); none of the other paired comparisons differed. Plumage scores of bright yearlings overlapped entirely with adult scores and bright

yearlings were treated as aggressively as bright adults by territory owners (Fig. 3); thus, they probably competed for territories in the same manner as adults. In contrast, dull yearlings were tolerated by adults (Fig. 3), which allowed them to settle nearby in high-quality habitat. The extreme tolerance shown by adult males toward some dull yearlings even suggested social bonds between them: during territorial settlement, some adults and dull yearlings were observed calmly perching or foraging together (Fig. 3). The failure of intermediate yearling males to settle on good territories suggests that they were too bright to be tolerated by adult males, but not aggressive enough to obtain territories by fighting.

Why would territorial adult males show such tolerance to dull males¹⁶, effectively permitting them to settle nearby? The possibility that yearling males dupe adult males by mimicking females (the “female mimicry hypothesis”¹⁰) was previously rejected experimentally⁹. Alternatively, the lack of aggression toward dull yearlings could reflect a form of male–male cooperation^{5,6} whereby both the dull yearling and adult male benefit. If females seek extra-pair matings on the basis of male plumage coloration^{17,18}, adult males might reduce their risk of cuckoldry by allowing less attractive (that is, dull) males to settle nearby¹⁹, and might also enhance their opportunities for obtaining extra-pair fertilizations^{16,20}.

Paternity analyses, using DNA fingerprinting, revealed a high level of extra-pair paternity in this socially monogamous species. Overall, 49% of nests ($n = 41$) contained at least one chick not sired by the resident male. These paternity analyses confirmed that adult males derive genetic benefits by having dull males as neighbours. First, dull males served as paternity buffers for adult males. The realized fitness of adult males was positively correlated with the proportion of their immediate territorial neighbours that were dull yearlings (Fig. 4a; Spearman rank correlation coefficient (r_s) = 0.40, $n = 22$, $P < 0.05$); this was due to a decreasing probability of fitness loss through extra-pair young (Fig. 4b: $r_s = -0.41$, $n = 22$, $P < 0.05$). In fact, from an adult male’s perspective, having all dull neighbours was genetically similar to having no immediate neighbours at all, a situation in which realized fitness tends to be high (Fig. 4a, right panel) and loss of paternity through extra-pair young tends to be low (Fig. 4b, right panel).

In addition, females paired with dull yearlings probably also provided a source of extra-pair paternity for adult males. This is because the plumage characteristics of a male influenced the likelihood that their nests contained extra-pair young: dull yearlings

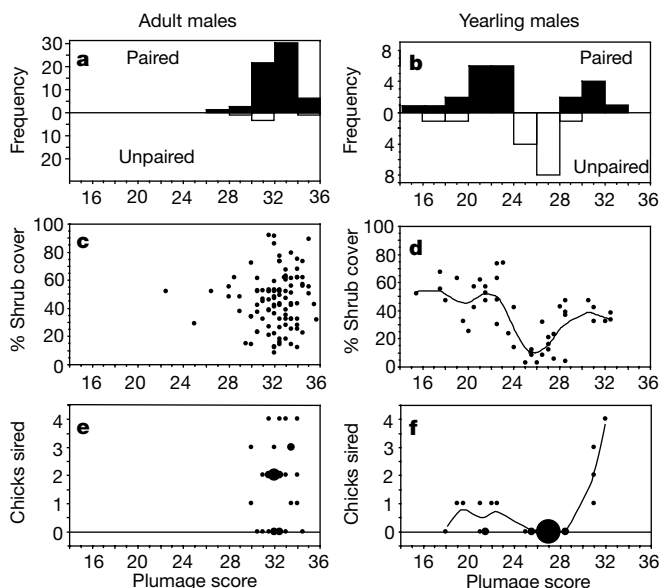


Figure 2 Social and genetic consequences of plumage variation in males. Left panels, adult males; right panels, yearling males. **a, b**, Distribution of plumage scores for males who obtained social mates (above) and those who did not (below). **c, d**, Relation between plumage score and territory quality. **e, f**, Relation between plumage colour and number of chicks sired in a male’s own nest, assessed by DNA fingerprinting. Unpaired males with no nests were assigned zeros. None of the comparisons for adult males were significant (all $P > 0.1$). All three comparisons for yearlings showed significant disruptive relations; univariate cubic splines were used to assess the form of phenotypic selection¹² (splines shown in **d** and **f**), and analysis with 100 bootstrapped splines¹² revealed significant minima ($P < 0.01$) in all three comparisons.

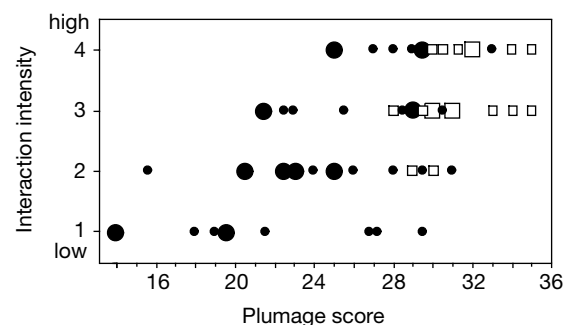


Figure 3 Relation between plumage coloration of an intruding male and the maximum intensity of aggressive interactions with adult male territory holders during territory settlement. Filled circles denote interactions with yearling intruders, open squares interactions with adult intruders. All points represent different pairs of interacting males: small symbols represent one interaction; larger symbols indicate two overlapping points. Aggression categories, in increasing levels of aggression, are (1) males perched less than 2 m apart with no visible sign of aggression; (2) males engaged in countersinging but no chases; (3) territory owner chased intruder, but no physical contact; and (4) territory owner chased intruder and physical contact observed. Aggression from adult male territory holders was positively correlated with plumage coloration of intruders ($r_s = 0.62$, $n = 59$, $P < 0.0001$).

(plumage score less than 24) were more than twice as likely as all brighter males to have at least one extra-pair nestling in their nests (87.5% of 8 nests versus 39% of 33 nests of brighter males; $G = 4.43$, $P < 0.05$). Consequently, dull males also had significantly more extra-pair young in their nests (1.5 ± 1.1 , $n = 8$) than brighter males (0.76 ± 1.1 , $n = 33$; Mann–Whitney $U = 191$, $P < 0.001$). If extra-pair matings primarily involve close neighbours²¹, adult males would increase their probability of siring extra-pair young by allowing dull yearlings to settle nearby. Thus, by tolerating dull yearlings, adult males shape the composition of their social neighbourhood to their own genetic advantage.

Dull yearlings also appear to benefit from their interactions with adult males, as they experience high pairing success (Fig. 2b). But given their high cuckoldry rates, do they actually increase their fitness through this interaction? Additional selection analysis on yearling plumage coloration, based on the number of chicks that males sire in their own nests, confirms that they do benefit. The significant minimum on the fitness surface (Fig. 2f, $P < 0.01$) indicates that dull yearlings sired more chicks than intermediate males.

Thus, settling in high-quality habitat increased the pairing success of the dull yearlings more than enough to compensate for the high levels of extra-pair paternity in their nests. This analysis also provides additional, even stronger evidence for disruptive selection on plumage, given that the number of chicks sired is a more accurate measure of fitness than pairing success. Dull yearlings also gained future benefits from settling in high-quality habitat: yearling territory quality was positively correlated with the quality of their territory the following year (correlation between percentage shrub cover on yearlings' territories and the percentage cover during the next year, $r = 0.71$, $n = 17$, $P < 0.05$). By demonstrating that both adult and dull yearling males benefited from their close proximity in high-quality habitat, our results provide rare evidence for sexually selected male–male cooperation^{5,6}.

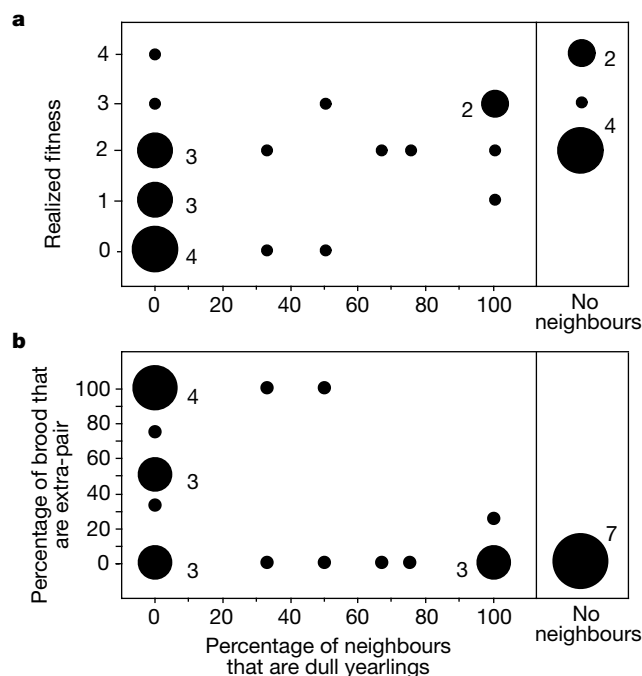


Figure 4 Dull yearling neighbours are paternity buffers for adult males. **a**, Relation between realized fitness in nests of adult males and the percentage of their immediate male neighbours (that is, those that shared a territory boundary) that were dull yearlings (plumage score less than 24). **b**, Relation between fitness loss through extra-pair young in the nest of adult males and the proportion of their male neighbours that were dull yearlings. Smallest points represent a single nest; sample sizes of nests indicated beside larger points. Panels on the right show data for adult males with isolated territories, sharing no territory boundary with another male.

The disruptive phenotypic selection on yearling plumage coloration thus resulted from plumage-dependent differences among males in their ability to acquire high quality territories, either involving cooperation or conflict with adult males. Extreme variation in yearling plumage was an important prerequisite, raising questions about the underlying proximate causes and maintenance of this variation. Two explanations are possible: the variation is largely genetic and is maintained as adaptive variation, or it is largely environmental and is maintained due to constraints^{2,10,11,22}. With respect to the former, strong disruptive selection can lead to equal-fitness alternative tactics¹⁴.

However, fitness comparisons reveal that the dullest yearling males (scoring below 24) had lower reproductive success (Fig. 2f; 0.56 ± 0.18 (s.e.m.) chicks sired in own nest, $n = 9$) than the brightest yearlings (2.5 ± 0.64 chicks sired, $n = 4$; Mann–Whitney $U = 33.5$, $P < 0.05$), a finding that is consistent with a constraint explanation. Moreover, the high frequency of males with intermediate plumage, despite selection against their phenotype, further suggests that the plumage variation is due largely to environmental factors (for example, hatching date, time of moult, condition during moult), and does not have a strong genetic basis. Finally, lazuli buntings breed in a wide variety of habitats⁷, and the pattern of selection may vary with habitat. For example, in habitats with extensive dense nesting cover, there is little competition among males for the unlimited high quality territories, and plumage coloration does not influence a yearling male's ability to acquire a good territory (B.E.L. and E.G., unpublished data). Thus, environmental factors may explain why many yearling males are unable to obtain an optimal plumage coloration. However, the behavioural flexibility to deal with this constraint mitigates some of the fitness losses; cooperation with adult males enables the dullest yearlings to make the best of a bad situation.

We found that complex patterns of mating success, in a species with an apparently simple mating system of social monogamy, derived ultimately from variation in territory quality, an ecological factor that mattered to females and led to striking spatial structure among males. Such spatial structuring may be widespread in other species that occur in variable habitats. Differences among species in variation in territory quality might also help explain currently puzzling variation in mating systems, such as differences in the frequency and distribution of extra-pair copulations. Historically, the study of mating systems focused on the ecological factors that shape social behaviour^{23,24}, but more recently emphasis has shifted away from ecology, to the social and particularly the genetic aspects of mating systems^{25,26}. Our findings suggest that some of the complexities of mating systems can best be unravelled by a combination of these approaches. □

Methods

Colour scoring

Males were captured in mist nets (Banding Permit no. 21830 to E.G.), and a standard series of photographs were taken showing all body parts. Adult and yearling males were aged by the colour of their primary wing covert feathers²⁷. Plumage scores were derived by summing scores from nine distinct areas on the head, breast, back and wings (details available at <http://biology.umd.edu/faculty/greene>, or as a colour figure provided on request from E.G.). Unlike other species whose plumage reflects strongly in the near-ultraviolet²⁸, the peak intensity of the blue feathers of lazuli buntings is at 500–575 nm in the visible portion of the spectrum (measured by reflectance spectrophotometry with an Ocean Optics S2000 fibre optic spectrometer with deuterium-halogen and tungsten illumination). Thus, our scoring does not miss brightness in the ultraviolet. This plumage scoring system is highly repeatable²⁹ by individuals (repeatability index, $r = 0.95$, $n = 24$ males), and among different individuals ($r = 0.93$, $n = 24$ males). The plumage of 18 males observed in aggressive interactions were scored in the field using a telescope. Plumage scores are very similar for birds scored from both photographs and through a telescope ($r = 0.91$, $n = 43$ males).

Territory quality

The percentage shrub cover on each male's territory was estimated in the field by averaging the percentage shrub cover estimated for four randomly placed 20 m × 20 m quadrats. At this study site, average territory size was 1.4 hectares (range 0.6–4.7 hectares, $n = 116$).

territories⁷). Thus, four quadrats sample about 11% of a territory of average size, and this was found to characterize the overall shrub cover well.

Aggressive interactions

Five-minute observations of interactions between territorial adult males and intruders were made during territory establishment during the 1994–1997 breeding seasons, and the maximum intensity of interaction was scored. Only one five-minute observation period was used to characterize levels of aggression between pairs of interacting males (that is, there was no pseudoreplication).

DNA fingerprinting

To assess parentage we used standard multilocus DNA fingerprints, based on restriction enzyme *HaeIII* and probes Per and Jeffreys 33.15 (following the protocols described in ref. 30). Autoradiographs were scored by eye. A chick was considered illegitimate if it had three or more unique bands and a bandsharing coefficient, D_i , of less than 0.35 with respect to the putative father. In all cases of extra-pair parentage, the male was excluded as the parent, never the female.

Statistical analyses

P values reported with splines indicate the number of times (out of 100 bootstrapped spline fits) that the resultant spline was not disruptive (that is, there was no clear minimum).

Received 7 July; accepted 2 August 2000.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1881).
2. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
3. Lande, R. Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
4. Endler, J. A. *Natural Selection in the Wild* (Princeton Univ. Press, Princeton, 1986).
5. McDonald, D. B. & Potts, W. K. Cooperative display and relatedness among males in a lek-mating bird. *Science* **266**, 1030–1032 (1994).
6. Noë, R. & Hammerstein, P. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav. Ecol. Sociobiol.* **35**, 1–11 (1994).
7. Greene, E., Muehler, V. R. & Davison, W. in *Birds of North America* (eds Poole, A. & Gill, F.) No. 232 (The Academy of Natural Sciences, Philadelphia & The American Ornithologists' Union, Washington DC, 1996).
8. Price, T. D. Sexual selection on body size, territory and plumage variables in a population of Darwin's Finches. *Evolution* **38**, 327–341 (1984).
9. Muehler, V. M., Greene, E. & Ratcliffe, L. Delayed plumage maturation in Lazuli buntings: tests of the female mimicry and status signalling hypotheses. *Behav. Ecol. Sociobiol.* **41**, 281–290 (1997).
10. Rohwer, S., Fretwell, S. D. & Niles, D. M. Delayed plumage maturation and the deceptive acquisition of resources. *Am. Nat.* **115**, 400–437 (1980).
11. Lyon, B. E. & Montgomerie, R. D. Delayed plumage maturation in passerine birds: reliable signaling by subordinate males? *Evolution* **40**, 605–615 (1986).
12. Schluter, D. Estimating the form of natural selection on a quantitative trait. *Evolution* **42**, 849–861 (1988).
13. Smith, T. B. Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes ostrinus*. *Nature* **363**, 618–620 (1993).
14. Gross, M. R. Disruptive selection for alternative life history strategies in salmon. *Nature* **313**, 47–48 (1985).
15. Sappington, T. W. & Taylor, O. R. Disruptive sexual selection in *Colias eurytheme* butterflies. *Proc. Natl Acad. Sci. USA* **87**, 6132–6135 (1990).
16. Morton, E. S., Forman, L. & Braun, M. Extra-pair fertilizations and the evolution of colonial breeding in purple martins. *Auk* **107**, 275–283 (1990).
17. Møller, A. P. & Birkhead, T. The evolution of plumage brightness in birds is related to extra-pair paternity. *Evolution* **48**, 1089–1100 (1994).
18. Yezerinac, S. M. & Weatherhead, P. J. Extra-pair mating, male plumage coloration and sexual selection in yellow warblers (*Dendroica petechia*). *Proc. R. Soc. Lond. B* **264**, 527–532 (1997).
19. Dugatkin, L. A. & Sargent, R. C. Male-male association patterns and female proximity in the guppy, *Poecilia reticulata*. *Behav. Ecol. Sociobiol.* **35**, 141–145 (1994).
20. Wagner, R. H. The pursuit of extra-pair copulations by female birds: a new hypothesis of colony formation. *J. Theor. Biol.* **163**, 333–346 (1993).
21. Gibbs, H. L. et al. Realized reproductive success of polygynous red-winged blackbirds revealed by DNA markers. *Science* **250**, 1394–1397 (1990).
22. Rohwer, S. & Butcher, G. S. Winter versus summer explanations of delayed plumage maturation in temperate passerine birds. *Am. Nat.* **131**, 556–572 (1988).
23. Orians, G. H. On the evolution of mating systems in birds and mammals. *Am. Nat.* **103**, 589–603 (1969).
24. Emlen, S. T. & Oring, L. W. Ecology, sexual selection and the evolution of animal mating systems. *Science* **197**, 215–223 (1977).
25. Birkhead, T. R. & Møller, A. P. *Sperm Competition in Birds: Evolutionary Causes and Consequences* (Academic, London, 1992).
26. Davies, N. B. Sexual conflict and the polygamy threshold. *Anim. Behav.* **38**, 226–234 (1989).
27. Young, B. E. Annual molts and interruption of the fall migration for molting in Lazuli buntings. *Condor* **93**, 236–250 (1991).
28. Hunt, S., Bennett, A., Cuthill, I. & Griffith, R. Blue tits are ultraviolet tits. *Proc. R. Soc. Lond. B* **265**, 451–455 (1998).
29. Lessells, C. M. & Boag, P. T. Unrepeatable repeatabilities: a common mistake. *Auk* **104**, 116–121 (1987).
30. Otter, K., Ratcliffe, L., Michaud, D. & Boag, P. T. Do female black-capped chickadees prefer high-ranking males as extra-pair partners? *Behav. Ecol. Sociobiol.* **43**, 25–36 (1998).

Acknowledgements

S. Cosh and T. Rooneem ran the DNA fingerprinting. We thank A. Keyser and G. Hill for

assistance in measuring reflectance of bunting plumage, and J. Elliott and B. Sinervo for assistance with the splines. We thank K. Bright, A. Chaine, K. Dial, D. Emlen, A. Greene, J. Jolivet, R. Hutto, S. Jones, R. Montgomerie, H. Powell, K. Short, B. Sinervo, B. Walker and K. Wasson for comments on this manuscript. We were assisted in the field by A. Agather, J. Carlson, W. Davison, R. Domenech, A. Edmonds, K. Grey, D. Gryskiewicz, J. Haskell, Q. Hodgson, K. Horst, K. Karwacky, L. Keeton, A.M. Lareau, J. Laws, J. Lee, L. Leroux, J. Lloyd, N. Marlenee, M. Miller, C. Minch, M. Miyai, A. Rapone, T. Redman, C. Richardson, J. Roach, J. Root, R. Sacco, R. Scholl, Y. Tamanda, A. Tomon, L. Whitney, B. Winter, K. Wood and J. York. This research was supported by the US Fish and Wildlife Service (to E.G.), Kananaskis Field Stations, University of Calgary (to B.E.L.), NSERC Collaborative Grant and NSERC Research Grants (to L.R. and P.T.B.), and an Ontario Graduate Scholarship, Queen's University School of Graduate Studies, American Museum of Natural History, Sigma Xi, Society of Canadian Ornithologists (to V.R.M.).

Correspondence and requests for materials should be addressed to E.G.
(e-mail: egreene@selway.umd.edu).

An area for vergence eye movement in primate frontal cortex

Paul D. Gamlin & Kyunghye Yoon

Vision Science Research Center, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

To view objects at different distances, humans rely on vergence eye movements to appropriately converge or diverge the eyes and on ocular accommodation to focus the object^{1,2}. Despite the importance of these coordinated eye movements (the 'near response') very little is known about the role of the cerebral cortex in their control. As near-response neurons exist within the nucleus reticularis tegmenti pontis³, which receives input from the frontal eye field region of frontal cortex^{4–6}, and this cortical region is known to be involved in saccadic^{7–9} and smooth-pursuit eye movements^{10–12}, we propose that a nearby region might play a role in vergence and ocular accommodation. Here we provide evidence from rhesus monkeys that a region of frontal cortex located immediately anterior to the saccade-related frontal eye field region is involved in vergence and ocular accommodation, and in the sensorimotor transformations required for these eye movements. We conclude that the macaque frontal cortex is involved in the control of all voluntary eye movements, and suggest that the definition of the frontal eye fields should be expanded to include this region.

Over the past decade, there has been significant progress in elucidating the anatomy and physiology of the subcortical pathways involved in the control of vergence eye movements and ocular accommodation in non-human primates^{13–15}. However, the areas of cerebral cortex involved in controlling these eye movements have remained elusive. On the basis of previous studies^{3–12}, we hypothesized that the frontal eye fields or an adjacent region within the frontal cortex might play a role in controlling vergence and ocular accommodation. To investigate this possibility, we recorded from the frontal cortex of two, trained, rhesus monkeys while they performed vergence, saccadic and smooth-pursuit eye movements. We encountered saccade-related neurons within the anterior bank of the arcuate sulcus⁸. In addition, in the prearcuate cortex rostral to the saccade-related region, we recorded from 34 neurons (22 from one animal and 12 from the other) displaying activity that was significantly correlated with both vergence and accommodation. Of these neurons, 25 increased their activity for near viewing, and 9 increased their activity for far viewing. As vergence and accommodation are highly correlated under normal viewing conditions, we related neuronal activity to the vergence eye movements. But previous studies^{16,17} suggest that some of the 34 neurons that we identified in this study are likely to be related to ocular accommodation as well

as to vergence eye movements. We identified 20 neurons that displayed phasic activity correlated with vergence dynamics, 9 neurons that displayed tonic activity correlated with vergence angle, and 5 neurons that displayed both phasic and tonic activity. Figure 1a–d shows an example of the behaviour of one of the convergence-related phasic neurons during vergence and saccadic eye movements. During convergence eye movements—elicited by targets stepped in depth—the activity of the neuron correlates with vergence velocity ($1.0 \text{ spikes s}^{-1} \text{ per degrees s}^{-1}$, $r = 0.57$, $P < 0.001$), and leads it by approximately 70 ms (Fig. 1a). Overall, the activity of these neurons led vergence eye movements by 28 ms (s.e.m., 7.5 ms; $n = 22$).

To investigate the possibility that the neural activity observed during such stepped target vergence trials was related not to the vergence movement but to the retinal disparity of the target, we examined the behaviour of neurons during vergence movements elicited by targets ramped in depth (see Methods, Behavioural tasks). As shown in Fig. 1b, the relationship between neural activity and vergence velocity under this condition is comparable to that seen in Fig. 1a, and thus good evidence that the behaviour of this neuron is more closely related to the movement than to the retinal

disparity of the target that elicited it. In addition, this neuron was tested during monocular accommodative vergence trials in which no retinal disparity was present, and increases in vergence angle were elicited through the blur-driven accommodative system. We found that the activity of this neuron under these conditions showed essentially the same relationship to vergence velocity ($0.9 \text{ spikes s}^{-1} \text{ per degrees s}^{-1}$, $r = 0.6$, $P < 0.001$) as it did when the eye movement was elicited by stepped changes in target distance during binocular viewing. This is further evidence that the behaviour of the neuron was more closely related to the motor output than to the sensory stimuli eliciting the movement. Such control trials were used to classify all neurons into the categories reported in this study.

We further examined the possibility that neurons displayed visual responses. To do this, we adapted a blanking model that has been used for this purpose in both cortical and subcortical studies of the neural control of smooth pursuit^{18,19}. More specifically, the behaviour of neurons were examined during binocular vergence tracking of a target that moved sinusoidally in depth and was briefly extinguished (250 ms) during the tracking period (Fig. 1c). The neuron shown in this figure maintained its activity during this blanking period despite the absence of visual stimulation, and

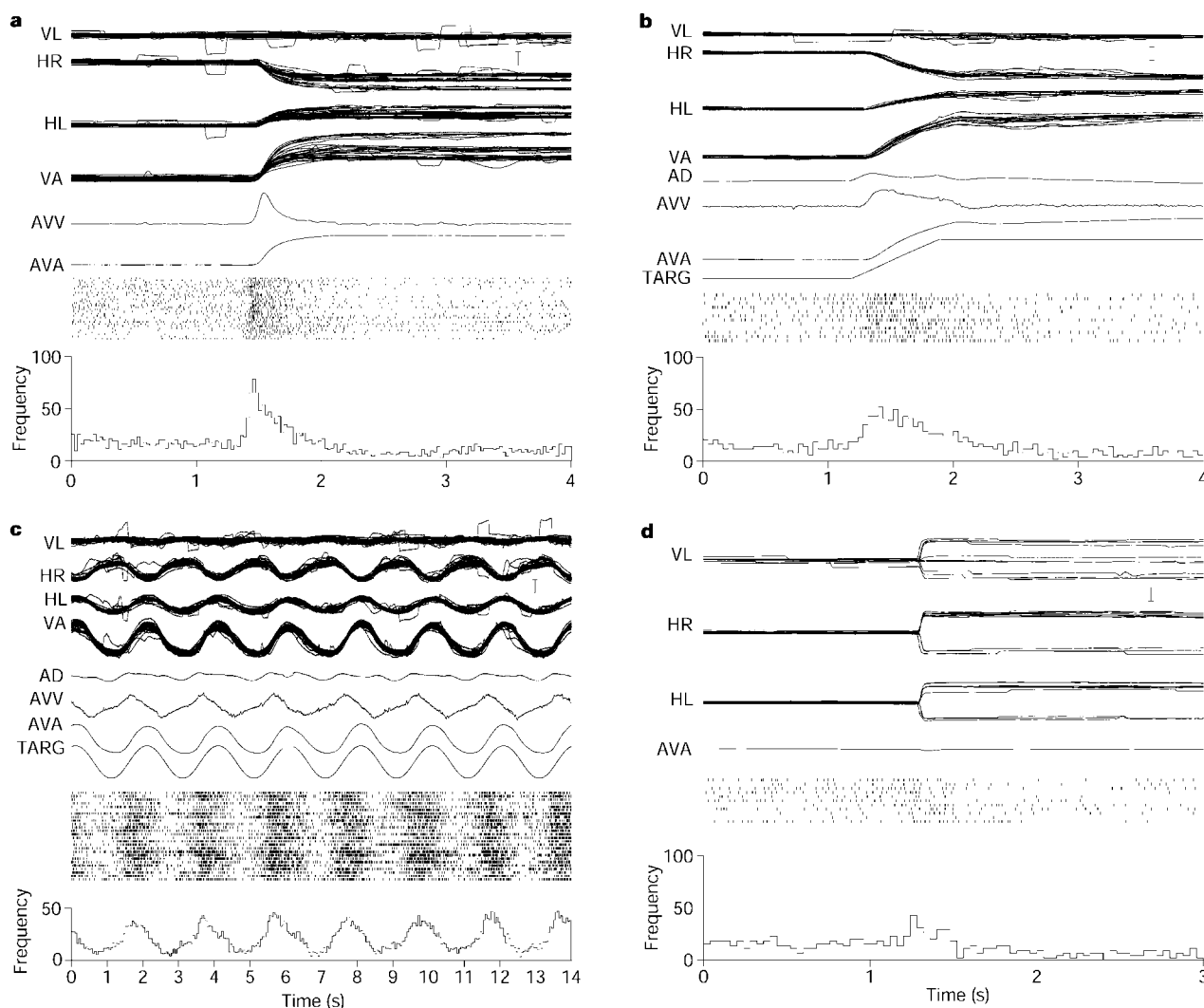


Figure 1 Response properties of a prearcuate vergence-related neuron. In **a** and **d**, rasters showing unit activity are aligned on eye movement onset. In **b** and **c**, they are aligned on target motion. Peristimulus time histograms at the base of each plot represent the average of the rasters. **a**, Neuronal response during vergence eye movements to a target stepped in depth. **b**, As **a** but for a target ramped in depth. **c**, Response of the neuron during sinusoidally modulated vergence responses in which the target was briefly

extinguished (blank in TARG trace). **d**, Response of the neuron for saccadic eye movements. AD, average disparity; AVA, average vergence angle; AVV, average vergence velocity; HL, horizontal left eye position; HR, horizontal right eye position; TARG, target position; VA, vergence angle; VL, vertical left eye position. Scale bar, 4 degrees, and 20 degrees s^{-1} .

therefore did not appear to display any visual signal. This result was similar for 20 neurons tested. Further, the activity of this neuron was not modulated during saccadic eye movements (Fig. 1d). Overall, of the 34 vergence-related neurons that were tested, only six showed significant alteration in their activity during conjugate eye movements.

The behaviour of 25 neurons was investigated during trials in which a target was presented, but a motor response was either delayed or not required (see Methods, Behavioural tasks). During these trials, most neurons, including that shown in Fig. 1, responded neither to the appearance of the visual target nor during the time period between its appearance and the movement. However, six neurons showed clear modulation of their activity under these conditions. An example of one of these visuo-vergence neurons is shown in Fig. 2. This neuron was initially identified as being vergence-related on the basis of its response during convergence eye movements elicited by stepped vergence targets (Fig. 2a). However, we found that when the animal was required to make delayed movements to near targets in the contralateral hemifield, the neuron significantly increased its activity in response to the appearance of the target—well in advance of the movement (Fig. 2b). In contrast, the activity of this neuron was modulated only weakly during the delay period for movements to far targets within the contralateral hemifield (Fig. 2c), or near targets within the ipsilateral hemifield (Fig. 2d). Thus, this neuron behaved as though it had both movement and visual receptive fields tuned for near space within the contralateral hemifield. Overall, the six neurons that responded during these trials behaved as though they have both movement and visual receptive fields tuned in depth within the contralateral hemifield. Further study will be required to determine whether this apparent visual response reflects a sensory component related to the visual characteristics of the target, or instead reflects movement-related activity such as that reported for some saccadic neurons^{20,21}.

Electrical microstimulation was performed at six sites where vergence-related neurons were encountered. At four of these sites, using stimulation currents of 40 μ A, we observed short-latency, stimulus-locked increases in convergence and ocular accommodation. Also, for 500–800 ms following cessation of electrical microstimulation, both the vergence and accommodation traces displayed damped, oscillatory movements that appeared to result from interactions between visual feedback and the effects of microstimulation. An example of the effect of microstimulation at the site of a convergence-related neuron is shown in Fig. 3. The animal binocularly fixated a target at an effective distance of 1 m during the trial. At this and the other effective sites, using a 500-Hz, 40-ms stimulus train, convergence was elicited with a latency of 34 ± 8 (s.d.) ms and reached a peak velocity of 5.1 ± 2.2 (s.d.) degrees s^{-1} . Accommodative responses were observed at a latency of 98 ± 11 (s.d.) ms. The observed latencies are consistent with the latency of 30 ± 14 (s.d.) ms reported for vergence movements elicited by microstimulation of the midbrain near-response region¹⁶ and with the latency of 75 ± 9 (s.d.) ms reported for accommodation elicited by microstimulation of the Edinger–Westphal nucleus²².

On the basis of the localization of our recording sites, we generated a composite diagram (Fig. 4) showing the approximate extent of the region of frontal cortex related to vergence and accommodation. This prearcuate region lies anterior to the saccade-related region of the frontal eye fields, and includes the surface cortex between the arcuate sulcus and the posterior pole of the principal sulcus. However, the boundaries shown in Fig. 4 should be considered only approximate. In both animals, once vergence-related neurons were encountered, we proceeded to record from a relatively localized area, and within this area it was possible to place only a relatively few marking lesions. Thus, a complete survey of the extent of this area remains to be conducted.

The results that we report here are, to our knowledge, the first unequivocal demonstration in non-human primates of a role for

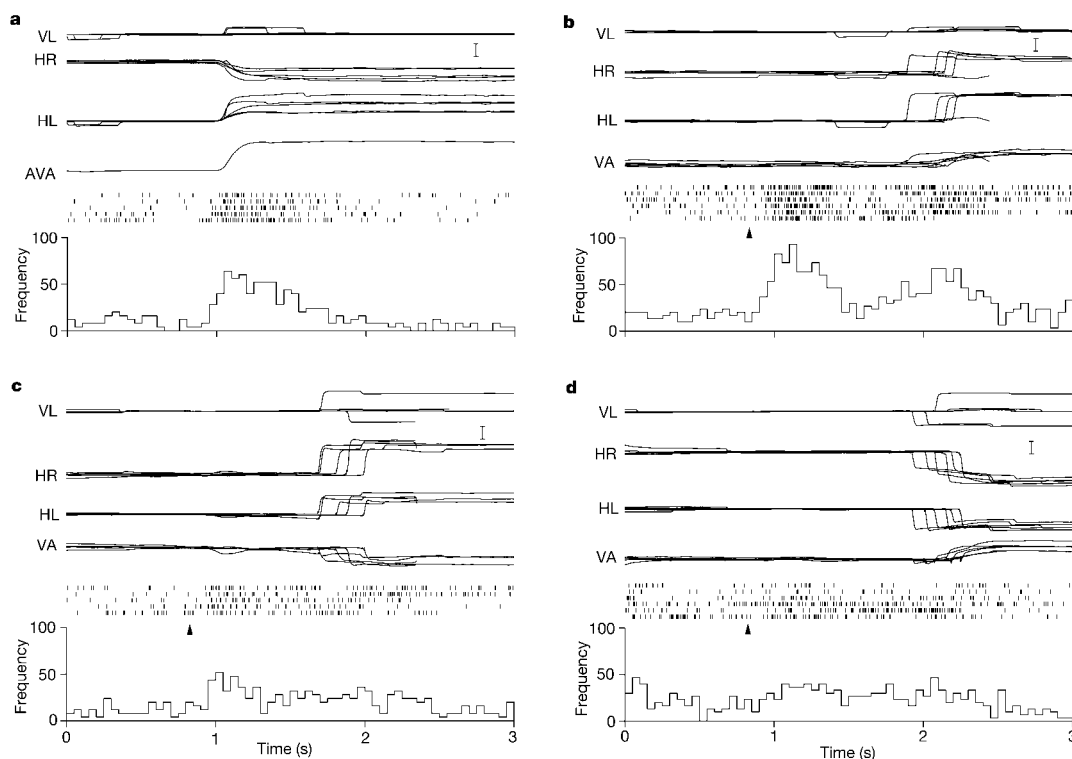


Figure 2 Response properties of a prearcuate visuo-vergence neuron. In **a**, the raster showing unit activity is aligned on eye movement onset. In **b**, **c** and **d**, the rasters are aligned on target appearance (indicated by the arrowhead). **a**, Neural response during convergence eye movements elicited by stepped increases in vergence demand.

b, Response to the appearance of a near target ($+4^\circ$ disparity) in the contralateral hemifield. **c**, Response to the appearance of a far target (-4° disparity) in the contralateral hemifield. **d**, Response to the appearance of a near target ($+4^\circ$ disparity) in the ipsilateral hemifield. Scale bar, 4 degrees.

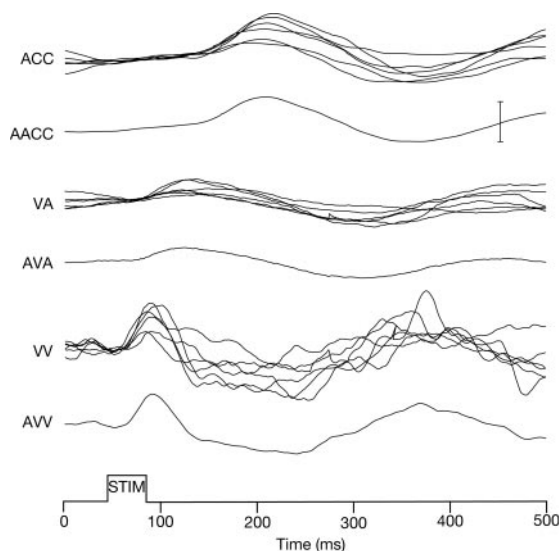


Figure 3 Effects of electrical microstimulation at the site of a convergence-related neuron. Biphasic electrical microstimulation elicited increases in vergence and accommodation. Stimulation parameters; 500 Hz, 40 ms, 40 μ A. AACC, average accommodation; ACC, accommodation; AVA, average vergence angle; AVV, average vergence velocity; VA, vergence angle; STIM, stimulation; WV, vergence velocity. Scale bar, 0.5 degrees, 0.5 diopters, and 5 degrees s^{-1} .

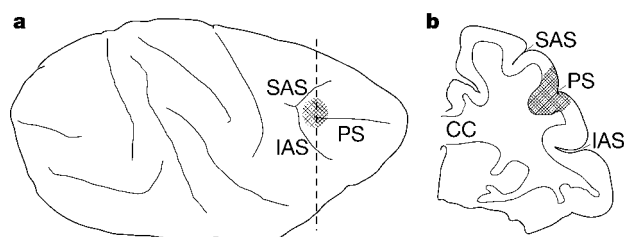


Figure 4 Location of near- and far-response neurons. The location of these neurons is shown as the shaded area in a side view of the rhesus cortical surface (**a**) and a coronal section (**b**) at the level indicated by the dashed line in **a**. CC, corpus callosum; IAS, inferior arcuate sulcus; PS, principal sulcus; SAS, superior arcuate sulcus.

the frontal cortex in both vergence and ocular accommodation. Thus, the arcuate region of frontal cortex contains areas specialized for the control of all classes of voluntary eye movements—saccades, smooth pursuit and vergence. This suggests that the boundaries of the primate frontal eye fields as defined at present should be expanded to include this vergence and ocular accommodation region, which lies anterior to the saccade-related region. $\square$

Methods

Animals

Two juvenile rhesus monkeys (*Macaca mulatta*) were used in this study. All experimental procedures were approved by the UAB Institutional Animal Care and Use Committee and complied with the USPHS Policy on Humane Care and Use of Laboratory Animals. All the surgical procedures described were performed under sterile conditions using either halothane or sodium pentobarbital anesthesia. Animals received analgesics (Numorphan, 0.15 mg kg^{-1} or Buprenex, 0.01 mg kg^{-1}) after surgery to minimize pain.

Behavioural tasks

The optical system and mechanisms for target display and movement have been described previously²³. Animals were trained to make vergence eye movements in response both to binocular and monocular targets that were either stepped or ramped towards or away from the animal. The binocular ramped trials were used to investigate the possibility that the neural activity observed during the stepped target vergence trials was related to the retinal disparity of the target and not to the vergence movement. When changes in vergence angle are elicited by step changes in target vergence, the retinal disparity of the target is initially equal to the target step. This disparity only starts to diminish when the vergence movement begins, and is minimized only at the completion of the movement.

There are neurons within frontal eye fields²⁴ and other visual cortical areas²⁵ that are sensitive to such disparities. By using a trial in which target vergence is ramped and not stepped, the retinal disparity of the target is minimized throughout the trial, and a neuron sensitive to the disparities present in the stepped target trial would not be expected to be active during such a trial. In contrast, vergence velocity is relatively unaffected during such ramped trials, and the behaviour of a vergence-related neuron should be similar under the two conditions. In addition, animals were trained to track targets moving sinusoidally in depth and to maintain accurate tracking during brief periods (≤ 300 ms) of target blanking. Animals were trained on saccadic and smooth-pursuit eye-movement tasks, and on visual probe and delayed-movement tasks. In the visual probe trial, the animal was required to maintain fixation of a central target during presentation of a second 'probe' stimulus. The optical system used in this study allowed both the horizontal and vertical location and disparity of this stimulus to be varied systematically. In the delayed-movement trial, which can also be considered as a "target overlap" trial²⁶, the animal was required to initially fixate a central target. Then, after a variable delay, a second target was presented while the animal maintained fixation. Both the horizontal and vertical location and disparity of this second target would be varied systematically. After a variable delay, the initial fixation point was extinguished and the animal was required to look to the second target.

Recording and microstimulation procedures

For eye movement recording, scleral search coils were implanted under the conjunctiva of each eye²⁷, and the horizontal and vertical gains of each eye were calibrated independently at the beginning of each recording session. Accommodation of the left eye was measured using a continuous recording, infrared optometer. Using a Kopf microdrive, parylene-insulated tungsten microelectrodes were advanced through a premeasured 26-gauge guide cannula contacting the dura. Unit activity was filtered sharply above 5 kHz, and the occurrence of a spike was detected with a window discriminator. As needed, the microelectrode could be switched to electrical microstimulation, which was produced using biphasic stimuli of 0.2 ms total duration at a frequency of 200–500 Hz generated by a Grass S88 stimulator and two stimulation isolation units. Microstimulation currents were generally 40 μ A and never exceeded 60 μ A. The positions of the right eye, left eye and accommodation were sampled at 1 kHz, while concurrent unit activity was sampled to the nearest 0.1 ms. All samples were stored on computer disk for later analysis.

Data analysis

Individual trials showing relevant eye movement traces and concurrent single unit activity were displayed and analysed off-line by computer. To determine the relationship of the neuron to the movement, plots of firing rate as a function of either vergence angle or velocity were produced, and correlation coefficients and regression parameters were determined for each plot as has been previously described^{23,28}. In addition, for visual probe or delayed movement trials, to determine if there were significant changes in the activity of the neuron during the trial, mean firing rates were measured for a fixed time period (500 ms) before and after the appearance of the target, and *t*-tests performed.

Verification of recording sites

For each animal, electrolytic lesions (30 μ A anodal; 20 s) were placed at the location of a few recording sites during the last two weeks of the study and, in one animal, a microinjection of the anatomical tracer, wheat germ agglutinin-horseradish peroxidase (WGA-HRP), was made near the centre of the region from which neurons had been recorded. Subsequently, standard histological procedures were used to localize both the marking lesions and the WGA-HRP injection site.

Received 15 May; accepted 7 September 2000.

- Rashbass, C. & Westheimer, G. Disjunctive eye movements. *J. Physiol.* **159**, 339–360 (1961).
- Schor, C. M. A dynamic model of cross-coupling between accommodation and convergence: Simulations of step and frequency responses. *Optom. Vis. Sci.* **69**, 258–269 (1992).
- Gamlin, P. D. R. & Clarke, R. J. Single-unit activity in the nucleus reticularis tegmenti pontis related to vergence and ocular accommodation. *J. Neurophysiol.* **73**, 2115–2119 (1995).
- Leichnetz, G. R., Smith, D. J. & Spencer, R. F. Cortical projections to the paramedian tegmental and basilar pons in the monkey. *J. Comp. Neurol.* **228**, 388–408 (1984).
- Huerta, M. F., Krubitzer, L. H. & Kaas, J. H. Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. Subcortical connections. *J. Comp. Neurol.* **253**, 415–439 (1986).
- Stanton, G. B., Goldberg, M. E. & Bruce, C. J. Frontal eye field efferents in the macaque monkey: I. Subcortical pathways and topography of striatal and thalamic terminal fields. *J. Comp. Neurol.* **271**, 473–492 (1988).
- Bizzi, E. Discharge of frontal eye field neurons during saccadic and following eye movements in unanesthetized monkeys. *Exp. Brain Res.* **6**, 69–80 (1968).
- Bruce, C. J. & Goldberg, M. E. Primate frontal eye fields. I. Single neurons discharging before saccades. *J. Neurophysiol.* **53**, 603–635 (1985).
- Dias, E. C., Kiesau, M. & Segraves, M. A. Acute activation and inactivation of macaque frontal eye field with GABA-related drugs. *J. Neurophysiol.* **74**, 2744–2748 (1995).
- Lynch, J. C. Frontal eye field lesions in monkeys disrupt visual pursuit. *Exp. Brain Res.* **68**, 437–441 (1987).
- MacAvoy, M. G., Gottlieb, J. P. & Bruce, C. J. Smooth-pursuit eye movement representation in the primate frontal eye field. *Cereb. Cortex* **1**, 95–102 (1991).
- Gottlieb, J. P., MacAvoy, M. G. & Bruce, C. J. Neural responses related to smooth-pursuit eye movements and their correspondence with electrically elicited smooth eye movements in the primate frontal eye field. *J. Neurophysiol.* **72**, 1634–1653 (1994).
- Leigh, R. J. & Zee, D. S. *The Neurology of Eye Movements* (Davis, Philadelphia, 1999).
- Gamlin, P. D. R., Yoon, K. & Zhang, H. The role of cerebello-ponto-cerebellar pathways in the control of

- vergence eye movements. *Eye* **10**, 167–171 (1996).
15. Gamlin, P. D. Subcortical neural circuits for ocular accommodation and vergence in primates. *Ophthalmic. Physiol. Opt.* **19**, 81–89 (1999).
 16. Judge, S. J. & Cumming, B. G. Neurons in the monkey midbrain with activity related to vergence eye movement and accommodation. *J. Neurophysiol.* **55**, 915–930 (1986).
 17. Zhang, Y., Mays, L. E. & Gamlin, P. D. R. Characteristics of near response cells projecting to the oculomotor nucleus. *J. Neurophysiol.* **67**, 944–960 (1992).
 18. Newsome, W. T., Wurtz, R. H. & Komatsu, H. Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *J. Neurophysiol.* **60**, 604–620 (1988).
 19. Mustari, M. J., Fuchs, A. F. & Wallman, J. Response properties of dorsolateral pontine units during smooth pursuit in the rhesus macaque. *J. Neurophysiol.* **60**, 664–686 (1988).
 20. Mays, L. E. & Sparks, D. L. Dissociation of visual and saccade-related responses in superior colliculus neurons. *J. Neurophysiol.* **42**, 207–232 (1980).
 21. Gnadt, J. W. & Andersen, R. A. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* **70**, 216–220 (1988).
 22. Gamlin, P. D. R., Zhang, Y., Clendaniel, R. A. & Mays, L. E. Behavior of identified Edinger-Westphal neurons during ocular accommodation. *J. Neurophysiol.* **72**, 2368–2382 (1994).
 23. Zhang, H. & Gamlin, P. D. R. Neurons in the posterior interposed nucleus of the cerebellum related to vergence and accommodation. I. Steady-state characteristics. *J. Neurophysiol.* **79**, 1255–1269 (1998).
 24. Ferraina, S., Pare, M. & Wurtz, R. H. Disparity sensitivity of frontal eye field neurons. *J. Neurophysiol.* **83**, 625–629 (2000).
 25. Fischer, B. & Poggio, G. F. Depth sensitivity of binocular cortical neurons of behaving monkeys. *Proc. R. Soc. Lond. B* **204**, 409–414 (1979).
 26. Munoz, D. P. & Wurtz, R. H. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *J. Neurophysiol.* **73**, 2313–2333 (1995).
 27. Judge, S. J., Richmond, B. S. & Chu, F. C. Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Res.* **20**, 535–538 (1980).
 28. Gamlin, P. D. R. & Mays, L. E. Dynamic properties of medial rectus motoneurons during vergence eye movements. *J. Neurophysiol.* **67**, 64–74 (1992).

Acknowledgements

We thank S. Hayley and K. Winston for assistance. This work was supported by the National Eye Institute.

Correspondence and requests for materials should be addressed to P.D.G. (e-mail: pgamlin@uab.edu).

The mammalian sodium channel BNC1 is required for normal touch sensation

Margaret P. Price[†], Gary R. Lewin[†], Sabrina L. McIlwraith[‡], Chun Cheng[†], Jinghui Xie[†], Paul A. Heppenstall[‡], Cheryl L. Stucky[‡], Anne G. Mannsfeldt[‡], Timothy J. Brennan[§], Heather A. Drummond[†], Jing Qiao[†], Christopher J. Benson[†], Deirdre E. Tarr[†], Ron F. Hrstkal^{||}, Baoli Yang^{||}, Roger A. Williamson^{||} & Michael J. Welsh[†]*

Howard Hughes Medical Institute, Departments of [†]Internal Medicine, [§]Anesthesia, ^{||}Obstetrics and Gynecology, and [¶]Physiology and Biophysics, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA
[‡]Growth Factors and Regeneration Group, Department of Neuroscience, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, D-13092 Berlin-Buch, Germany

* These authors contributed equally to this work

Of the vertebrate senses, touch is the least understood at the molecular level. The ion channels that form the core of the mechanosensory complex and confer touch sensitivity remain unknown^{1–3}. However, the similarity of the brain sodium channel 1 (BNC1)^{4–6} to nematode proteins involved in mechanotransduction indicated that it might be a part of such a mechanosensor^{7,8}. Here we show that disrupting the mouse BNC1 gene markedly reduces the sensitivity of a specific component of mechanosensation: low-threshold rapidly adapting mechanoreceptors. In rodent hairy skin these mechanoreceptors are excited by hair movement². Consistent with this function, we found BNC1 in the lanceolate nerve endings that lie adjacent to and surround the hair follicle⁹. Although BNC1 has been proposed to have a role in pH sensing^{10,11}, the acid-evoked current in cultured sensory neurons and the

response of acid-stimulated nociceptors were normal in BNC1 null mice. These data identify the BNC1 channel as essential for the normal detection of light touch and indicate that BNC1 may be a central component of a mechanosensory complex.

Stimulation of sensory receptors in the skin generates a variety of sensations, including touch and pain^{1,2}. However, these are the least understood vertebrate senses at the cellular and molecular level. Specifically, the ion channel receptors that convert mechanical stimuli into electrical ones in vertebrate nerve endings remain unknown³. A clue to their identity came from the discovery of MEC-4 and MEC-10 in a genetic screen of touch-insensitive *Caenorhabditis elegans* mutants^{7,8}. MEC-4 and MEC-10 belong to the DEG/ENaC family of proteins, which associate as homo- and heteromultimers to form voltage-insensitive Na⁺ channels¹². BNC1 (ref. 4) (also called MDEG (ref. 5), BNaC1 (ref. 6) and ASIC2 (ref. 10)) is one of several mammalian DEG/ENaC channels. The

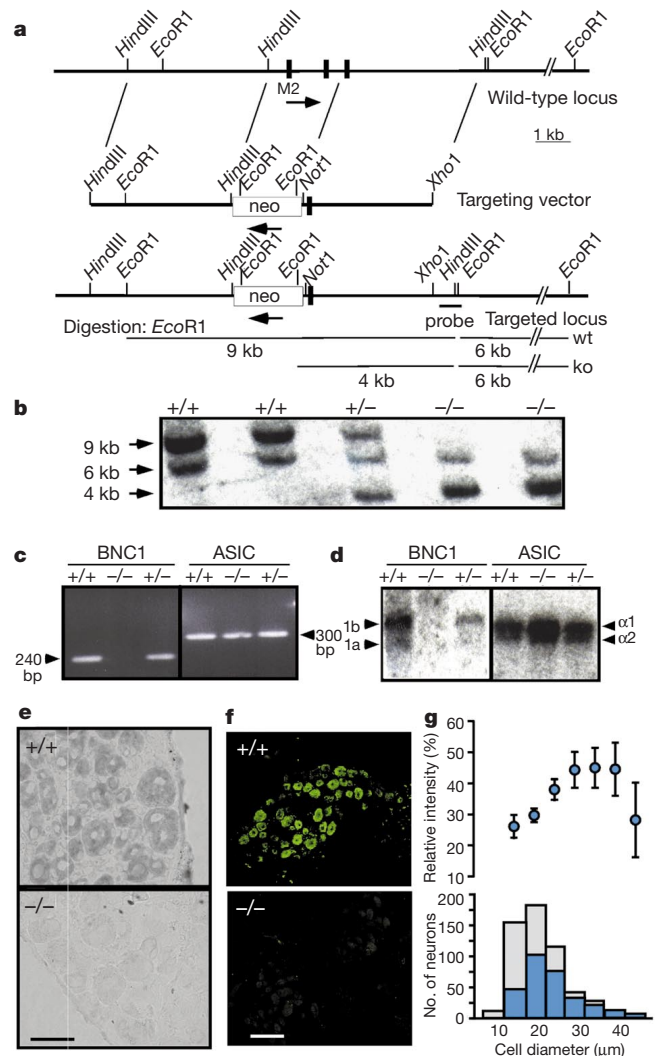


Figure 1 Disruption of BNC1 locus. **a**, Structure of wild-type BNC1 locus, targeting vector and resulting targeted locus. Insertion of neo cassette replaced region encoding M2 and exon 3' to this. Location of probe used for Southern blot is shown. Expected fragment sizes of wild-type (9 and 6 kb) and mutant (6 and 4 kb) alleles are indicated. **b**, Southern blot of EcoRI-digested genomic DNA. **c**, RT-PCR analysis of BNC1 and ASIC transcripts in brain RNA. BNC1 product 240 bp; ASIC product 300 bp. **d**, Northern blots of brain RNA. BNC1 probe, bases 1,351–1,536 BNC1a coding sequence, ASICα probe, bases 217–1,413. **e**, *In situ* hybridization using RNA probes to N terminus of BNC1b. Scale bar, 50 μm. **f**, Immunolocalization of BNC1 in sections of DRG tissue. Scale bar, 100 μm. **g**, Histogram (bottom) of the soma size of BNC1 positive and negative sensory neurons. Blue bars represent proportion of positive neurons in each diameter bin. Relative intensity (top) of antibody staining for positively stained neurons of indicated size.

fact that mutation of a specific residue in the nematode channels^{13,14} and in BNC1 (ref. 5) generates a constitutively active channel indicated that they might have functional similarity. Therefore, we hypothesized that BNC1 forms part of the mechanosensory complex in mammals. BNC1 might also be involved in acid-induced nociception^{10,11}; this hypothesis is based on the finding that when BNC1 is expressed in heterologous systems it is activated by acidic extracellular pH. To test these hypotheses, we disrupted the mouse *BNC1* gene.

We inserted a neo cassette to replace two exons encoding the second transmembrane domain (M2), the most highly conserved region of DEG/ENaC proteins and one that contributes to the channel pore¹². BNC1 has two splice variants, BNC1a and 1b (also called MDEG1 and MDEG2), which have different amino termini and first transmembrane domains^{4,5,15}. An earlier report indicated that BNC1b is expressed in dorsal root ganglion (DRG)¹⁵; our data indicate that BNC1b is the predominant transcript, although by

polymerase chain reaction with reverse transcription (RT-PCR) we detected both 1a and 1b transcripts in wild-type DRG (data not shown). Our targeting strategy ensured disruption of both splice variants (Fig. 1a). BNC1^{-/-} heterozygous and wild-type mice (Fig. 1b) were produced in the expected 1:2:1 Mendelian ratio, BNC1^{-/-} mice had normal appearance, growth, size, temperature, fertility and life span, and showed no obvious behavioural abnormalities. Using RT-PCR with primers from the targeted region and northern analysis using a probe common to both splice variants, we failed to detect transcripts in the brain of BNC1 null animals (Fig. 1c, d).

In DRG, most large sensory neurons are low-threshold mechanoreceptors that detect innocuous stimuli such as light touch, whereas most small neurons are nociceptive, detecting noxious mechanical, thermal or chemical stimuli^{2,3}. *In situ* hybridization and immunostaining detected transcripts and protein, respectively, in both large and small DRG neurons of wild-type animals but not

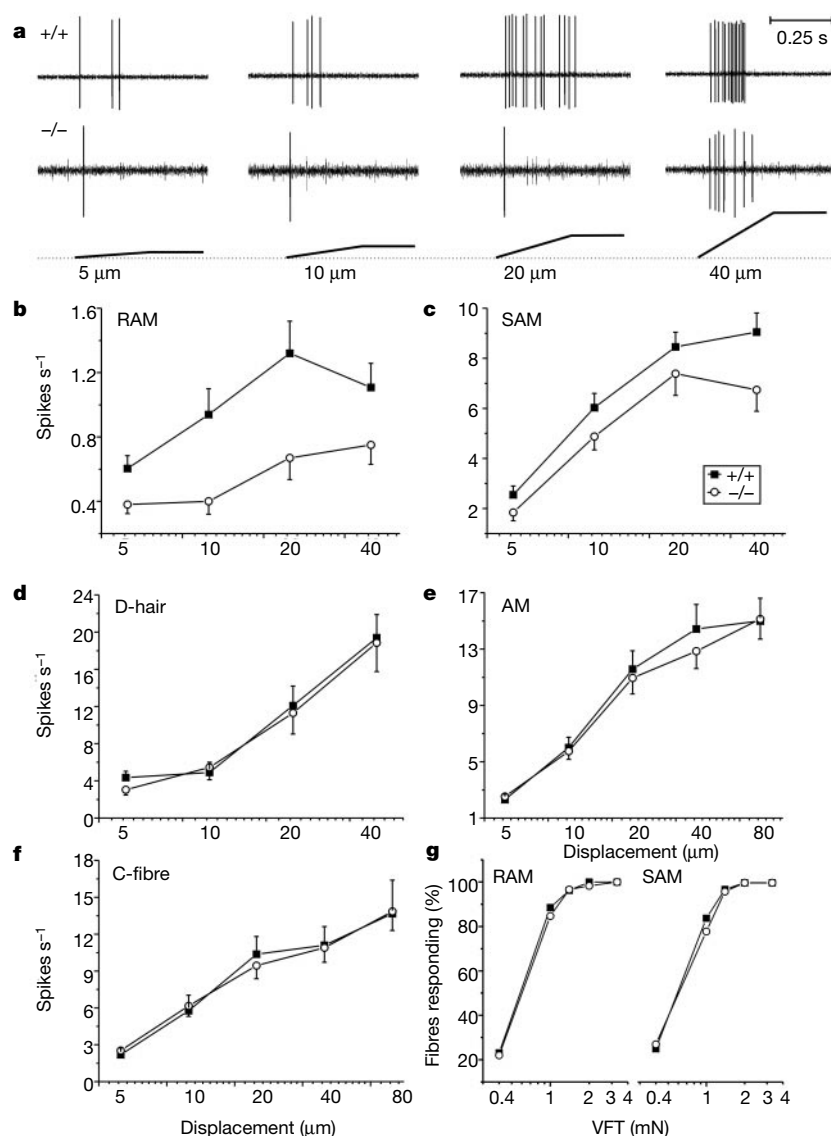


Figure 2 Mechanosensitivity of wild-type and BNC1 null mice. **a**, Example of RA mechanoreceptor from wild-type and BNC1 null mouse. **b**, Stimulus–response function of RA mechanoreceptor from wild-type (closed squares, $n = 48$) and null mice (open circles, $n = 27$), showing flattening of stimulus–response function of BNC1 null mice ($P < 0.005$ two-way analysis of variance (ANOVA)). At each stimulus point above $5 \mu\text{m}$, response in BNC1 null mice was significantly different from wild type, $P < 0.05$, unpaired t -test. **c**, Stimulus–response function of SA mechanoreceptor fibres. Null mice showed

significant decrease in sensitivity of SA mechanoreceptors ($n = 55$) compared with wild type ($n = 87$) ($P < 0.05$ two-way ANOVA). **d–f**, Stimulus–response function from slowly conducting D-hair afferents and myelinated and non-myelinated nociceptors ($n = 18–43$ for each fibre type and genotype). Frequency of firing was calculated over the entire 10 s of stimulus except in the case of D-hair afferents where only the first 2 s were used. **g**, Median force required to activate low threshold mechanoreceptors. VFT is von Frey threshold.

in null animals (Fig. 1e, f). In cultured DRG neurons both large and small neurons stained, but the largest neurons tended to show the most intense staining (Fig. 1g). Expression of BNC1 in small and large DRG sensory neurons is consistent with a role in both mechanosensation and acid-evoked nociception.

To learn whether BNC1 has a role in mechanosensation, we examined the response of single sensory neurons to mechanical stimuli using an *in vitro* skin-nerve preparation^{16,17}. Myelinated and unmyelinated single fibres were classified into rapidly adapting (RA) and slowly adapting (SA) low-threshold mechanoreceptors, D-hair receptors (D-hair), A-fibre mechano-nociceptors (AM) and C-fibre mechano-nociceptors¹⁶. We investigated the relationship between the strength of the stimulus and the neural response using a standard ascending series of mechanical stimuli. Increasing stimuli produced increasing numbers of action potentials in wild-type RA mechanoreceptors (Fig. 2a, b). In BNC1 null mice, we found a reduced sensitivity of the low-threshold RA mechanoreceptors; their stimulus-response function showed a reduced discharge frequency and flattening in the absence of the BNC1 channel. Thus, wild-type afferent neurons displayed their highest dynamic sensitivity between 5 and 20 μm displacement, whereas afferents from null mice showed a much reduced response in this range. We also found a smaller but significant shift in the stimulus-response function of SA mechanoreceptors (Fig. 2c). The proportion of fast conducting A-fibres classified as SA and RA was identical in BNC1^{-/-} (33% RA, 67% SA, $n = 82$) and BNC1^{+/+} mice (34% and 66%, $n = 132$). The stimulus-response functions of the rapidly adapting D-hair mechanoreceptors and the AM and C-fibre mechano-nociceptors were not affected by the absence of BNC1 (Fig. 2d–f). Using von Frey hairs, we found that the minimal force required to elicit action potentials was not different for RA mechanoreceptors, SA mechanoreceptors or other fibre types in BNC1 null mice (Fig. 2g and data not shown); however, this method may not be sensitive enough to detect a small shift in threshold, as it is limited by the number of hairs used in the lower range.

Sensory mechanotransduction is a two-step process¹⁸. First,

mechanically gated ion channels open, generating a graded receptor potential. Second, the membrane depolarization initiates action potentials. We tested whether the absence of the BNC1 channel might affect the second process. We current-clamped cultured large-diameter sensory neurons and studied cells with narrow spikes characteristic of low-threshold mechanoreceptors¹⁹. The amount of current injection required to initiate action potentials was identical in BNC1^{+/+} and BNC1^{-/-} neurons (Fig. 3a, b). Thus the reduced mechanosensitivity of low-threshold mechanoreceptors in BNC1 null mutants probably reflects a requirement for the channel in the generation of the mechanically gated receptor potential. However, the size of sensory nerve endings in skin precludes voltage-clamp recordings of receptor potentials *in situ*.

We also tested the role of BNC1 in acid-evoked responses. Acidic pH activates heterologously expressed BNC1a and the related neuronal channels, acid-sensing ion channel (ASIC) and dorsal root acid-sensing ion channel (DRASIC)¹⁰. Earlier studies showed acid-evoked Na⁺ currents in cultured large-diameter sensory neurons^{20,21}. Therefore, we tested the contribution of BNC1 to these currents. Whole-cell currents from large wild-type and null DRG neurons showed similar H⁺-gated transient and sustained currents (Fig. 3c). The pH sensitivity was the same, and the currents were blocked by amiloride (Fig. 3d, e). A similar fraction of wild-type (59%, $n = 29$) and null neurons (55%, $n = 33$) showed H⁺-gated currents. In addition, small-diameter neurons (<25 μm) showed the same inward current induced by pH5 in BNC1^{+/+} and BNC1^{-/-} neurons (7/16^{+/+} neurons, 210 ± 48 pA; 5/14^{-/-} neurons, 238 ± 101 pA). These data indicate that BNC1 makes little contribution to H⁺-gated currents in sensory DRG neurons.

Earlier studies have shown that acid applied to the receptive terminals of many polymodal nociceptors, or C-mechano-heat (C-MH) receptors, evokes a sustained discharge²². We found that the proportion of C-fibres classified as C-MH and C-mechano-nociceptors (C-M) was not different between wild-type (64%, $n = 39$) and BNC1 null mice (54%, $n = 45$, $p > 0.2$ χ^2 test). In both types of mouse, acid application (pH 5.0) induced a sustained discharge in

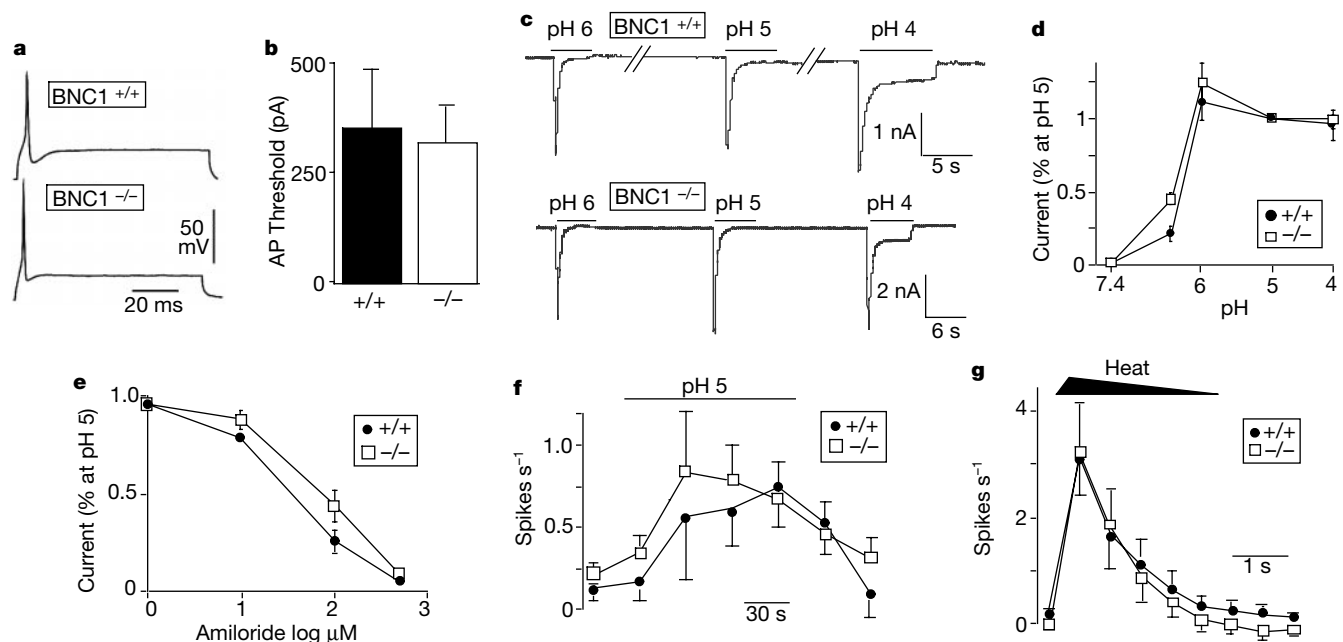


Figure 3 Response of sensory neurons to current injection, acidic pH and heat. **a**, Examples of current-evoked action potentials from large-diameter sensory neurons. **b**, Amount of current injection needed for large-diameter neurons (wild type $35 \pm 1.8 \mu\text{m}$ and null $33.5 \pm 1.7 \mu\text{m}$), mean $\pm$ s.e.m., $n = 5$ –8 per genotype. **c**, pH-dependent whole-cell currents. Voltage was -70 mV; downward deflection indicates inward current. Cells were $35.7 \pm 0.5 \mu\text{m}$ for wild-type and $35.8 \mu\text{m}$ for BNC1^{-/-} mice. **d**, Effect of pH on transient current, $n = 4$ –6 cells obtained from at least two mice. **e**, Effect of amiloride on current activated by pH 5; $n = 3$ –5 at each point. **f**, Response of single C-fibre nociceptors to pH 5. Data are from C-MH fibres ($n = 9$), the only fibres found to be consistently excited by low pH. **g**, Response of C-MH fibres to a brief noxious heat stimulus. Skin temperature was raised transiently to over 50°C and then rapidly fell ($n = 24$ –25 each genotype).

d, Effect of pH on transient current, $n = 4$ –6 cells obtained from at least two mice. **e**, Effect of amiloride on current activated by pH 5; $n = 3$ –5 at each point. **f**, Response of single C-fibre nociceptors to pH 5. Data are from C-MH fibres ($n = 9$), the only fibres found to be consistently excited by low pH. **g**, Response of C-MH fibres to a brief noxious heat stimulus. Skin temperature was raised transiently to over 50°C and then rapidly fell ($n = 24$ –25 each genotype).

around 35% of C-MH fibres. The time course and magnitude of response in C-MH fibres were identical (Fig. 3f). Consistent with earlier reports²², we saw no response to the pH 5 stimulus in C-M fibres or low-threshold mechanoreceptors from mice of either genotype. We also found that the response to a standardized noxious heat stimulus was the same in wild-type and null mice (Fig. 3g). These data indicate that sustained excitation of polymodal C-fibres by acidic pH does not require the BNC1 channel.

Rapidly adapting mechanoreceptors are stimulated by hair movement²; thus we hypothesized that BNC1 would be localized in nerve endings around hairs (Fig. 4, inset)⁹. We found BNC1 expressed in the palisades of lanceolate nerve endings. These fibres surround the hair follicle, running longitudinally to the shaft at and below the level of the sebaceous glands (Fig. 4a). Thus, they are well positioned to detect movement of the hair shaft⁹. In contrast, immunostaining was weaker and less consistent in the pilo-Ruffini

endings spiralling around the lanceolate nerve endings. Lanceolate endings did not stain in BNC1 null mice. The number and appearance of lanceolate fibres was similar in BNC1^{+/+} and BNC1^{-/-} animals as evaluated by blinded observers (Fig. 4b, c and Table 1).

The role of BNC1 in mechanosensation indicates evolutionary conservation of function within the DEG/ENaC channel family. A new *Drosophila* channel, NompC, unrelated to the DEG/ENaC family, has been described as necessary for mechanotransduction in ciliated mechanosensory neurons²³. It is unclear whether mechanotransduction in sensory hair cells requires channels from the DEG/ENaC family; of interest, the related α ENaC channel is not required for normal hair cell mechanotransduction²⁴.

Studies in humans and nonhuman primates indicate that the dynamic sensitivity of both RA and SA mechanoreceptors is critical for the perception and discrimination of touch sensation, especially that associated with hair movement and skin indentation^{1,2}. Our data show that BNC1 is specifically required for the sensitivity of these low-threshold mechanoreceptors. We speculate that mechanical deformation of sensory terminations, for example in hair follicles, may activate a channel complex that includes BNC1, thereby eliciting a depolarizing cation current that triggers action potentials. It is also clear, however, that RA and SA mechanoreceptors retain some mechanosensitivity in the absence of BNC1. Perhaps the mechanosensory channel complex is a heteromultimer composed of BNC1 and other DEG/ENaC subunits, and the absence of BNC1 precludes a normal graded response to mechanical stimuli. Of interest, the related DEG/ENaC channels β - and γ ENaC have been localized in rats to lanceolate nerve endings of the vibrissal follicle and Merkel cell and lamellated corpuscles in the foot pad^{25,26}. Moreover, it seems likely that additional proteins tether BNC1 to the extracellular matrix and intracellular cytoskeleton, thereby transmitting applied stresses to gate the channel^{7,8}. Our identification of a molecular component for a specific cutaneous sensory modality in vertebrates is an essential first step toward a molecular description of touch. □

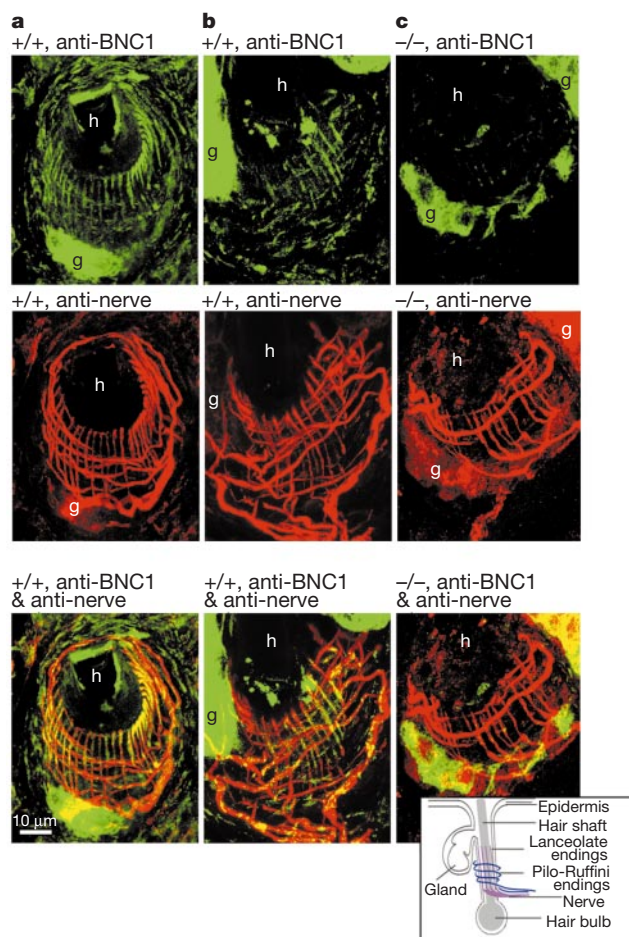


Figure 4 Immunostaining of BNC1 around guard hair follicles. **a**, Oblique sections through hair follicle of a wild-type mouse; micrographs were obtained by confocal microscopy with stacking of 12 images, each 0.8 μ m thick. Green indicates staining of BNC1 and red nerves. **b**, **c**, Complementary images from BNC1^{+/+} and BNC1^{-/-} hair follicles. Inset, neural network innervating guard hair follicle. h, hair follicle shaft; g, sebaceous gland. Sebaceous glands stain nonspecifically with primary and secondary antibodies.

Table 1 Evaluation of lanceolate fibres of hair follicles

	n	Hair diameter (μ m)	Number of lanceolate fibres per 16 μ m	Grade
BNC1 ^{+/+}	20	10.1 $\pm$ 0.9	6.5 $\pm$ 1.2	1.0 $\pm$ 0.6
BNC1 ^{-/-}	23	10.1 $\pm$ 0.9	7.2 $\pm$ 1.5	1.0 $\pm$ 0.8

Diameters of hairs and number of lanceolate fibres in 16 μ m were counted on guard hair follicles. Data are mean $\pm$ s.d. Appearance of fibres was graded from 2 (good morphology with regular distribution and density) to 0 (morphology irregular, fibres not distributed evenly, and low density). Observers were blinded to genotype.

Methods

Construction of targeting vector and generation of BNC1 mutant mice

We disrupted the BNC1 gene by homologous recombination using the P1339 LoxP-PGK-Neo vector (a gift from T. J. Ley, Washington University) and standard methods²⁷. We characterized mice generated from two different embryonic stem cell lines. For RT-PCR, we extracted mouse brain total RNA and used the following primers directed to the targeted region: BNC1 5' TCC GAG AAC ATT CTT GTT CTG GAT 3' and 5' GTT CTC ATC ATG GCT CCC TTC CTC 3' (240 base pair (bp) band); ASIC 5' CAG CTA GAG ATA TTG CAG GAC AAG 3' and 5' CCA CAC AGG CAA GTA CTT ATC TTG 3' (300 bp band).

Immunocytochemistry and *in situ* hybridization

We used a rabbit polyclonal anti-BNC1 antibody against amino acids 186–421 (numbering as in ref. 4) to recognize both BNC1a and BNC1b. DRG (10 μ m) and hairy skin (50 μ m) frozen sections were prepared as described²⁸. Sections were incubated with affinity-purified BNC1 antibody and Rhodamine Red X-conjugated donkey anti-rabbit IgG. In hairy skin, sections were also stained with a nerve cocktail consisting of sheep anti-neurofilament and sheep anti-PGP 9.5. We used a Biorad MRC1024 laser-scanning confocal microscope to examine samples.

For DRG neurons in culture, we used a rabbit polyclonal antibody (raised against amino acids 382–396, as in ref. 4) as described²⁸. Openlab software (Improvision) was used for quantification of the immunofluorescence. Ten individual cultures from five animals per genotype were done in parallel. Staining intensity of BNC1^{-/-} cultures was comparable to that of peptide controls. The mean fluorescence value for relative intensity was corrected for this background staining. For each cell, mean diameter was calculated and the proportion of cells positive for BNC1 was calculated for each bin.

In situ hybridization was performed with 10- μ m cryosections of lumbar DRG as described²⁸. Sections were hybridized with a Digoxigenin-labelled cRNA probe of 400 bp complementary to the N terminus of BNC1b; the same sequence was used to synthesize a 400-bp sense probe.

Culture of DRG neurons

DRG from 2–3-month-old mice were dissociated and cultured as described for rat²⁹ with the modification that ganglia were digested with papain (10 U ml⁻¹) and collagenase/dispase solution (1.67 μ g ml⁻¹) separately for 15 min at 37 °C. For the study of small-diameter neurons, we prepared cultures of adult DRG neurons as described²⁸. No nerve growth factor or other neurotrophins were added to the medium.

Whole-cell patch-clamp recording

DRG neurons were bathed in extracellular solution containing, in mM: 128 NaCl, 5 MgCl₂, 1.8 CaCl₂, 5.4 KCl, 5.6 glucose, 20 HEPES, pH 7.4. Acidic solutions were buffered by 10 mM MES and 10 mM HEPES. Pipettes contained, in mM: 100 KCl, 30 NaCl, 2 MgCl₂, 10 EGTA, 20 HEPES and 1 ATP. All cells exhibited depolarization-activated currents. Methods to assess pH-gated currents in small diameter neurons and membrane properties of isolated mechanoreceptors were as described³⁰. Under current-clamp conditions action potentials were evoked with current injection. Neurons with broad action potentials displaying a hump on the falling phase (a marker of nociceptor type neurons) were tested for pH 5-induced currents. The amount of current required to initiate action potentials in large diameter neurons with narrow spikes was also measured. Cells were used from at least four mice per genotype in each experimental series.

Single fibre recording

We used an *in vitro* skin/nerve preparation to record from functionally single primary afferents in micro-dissected teased filaments of the saphenous nerve as described^{16,17}. A standard ascending series of displacement stimuli were applied to the receptive field at 30-s intervals. Each displacement was maintained for 10 s, and neurons that maintained a discharge throughout the stimulus were characterized as SA mechanoreceptors. Those neurons that responded only at the beginning and end of the 10-s stimuli were classified as RA mechanoreceptors. Each stimulus-response function started at threshold as the probe was adjusted so that the first 5-μm displacement evoked spikes.

A group of C-fibre nociceptive neurons (axonal conduction velocity <1.0 m s⁻¹) was characterized with the same stimulus series (*n* = 46). C-fibres were not divided into C-MH and C-M fibres, as these two groups have essentially identical mechanosensitivity¹⁶. A separate series of experiments were carried out to characterize the heat and pH sensitivity of nociceptive C-fibres (80 C-fibres tested)²². A small metal ring isolated the receptive field and a heated Ringer solution was rapidly applied. The measured intradermal temperature increased to over 49 °C for about 3 s; this stimulus activates all heat-sensitive nociceptors. One min after the heat stimulus, the solution was replaced with oxygenated Ringer solution with pH 5.0 for exactly 2 min. Raw electrophysiological data were collected with a Powerlab 4.0 system (AD Instruments) and spikes discriminated off line with the spike histogram extension of the software.

Received 19 July; accepted 4 September 2000.

- Johnson, K. O. & Hsiao, S. S. Neural mechanisms of tactual form and texture perception. *Annu. Rev. Neurosci.* **15**, 227–250 (1992).
- Perl, E. R. in *Sensory Neuron: Diversity, Development, and Plasticity* (ed. Scott, S. A.) 3–23 (Oxford Univ. Press, Oxford, 1992).
- Caterina, M. J. & Julius, D. Sense and specificity: a molecular identity for nociceptors. *Curr. Opin. Neurobiol.* **9**, 525–530 (1999).
- Price, M. P., Snyder, P. M. & Welsh, M. J. Cloning and expression of a novel human brain Na⁺ channel. *J. Biol. Chem.* **271**, 7879–7882 (1996).
- Waldmann, R., Champigny, G., Voilley, N., Lauritzen, I. & Lazdunski, M. The mammalian degenerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in *Caenorhabditis elegans*. *J. Biol. Chem.* **271**, 10433–10436 (1996).
- García-Añoveros, J., Derfler, B., Neville-Golden, J., Hyman, B. T. & Corey, D. P. BNaC1 and BNaC2 constitute a new family of human neuronal sodium channels related to degenerins and epithelial sodium channels. *Proc. Natl Acad. Sci. USA* **94**, 1459–1464 (1997).
- Tavernarakis, N. & Driscoll, M. Molecular modeling of mechanotransduction in the nematode *Caenorhabditis elegans*. *Annu. Rev. Physiol.* **59**, 659–689 (1997).
- García-Añoveros, J. & Corey, D. P. The molecules of mechanosensation. *Annu. Rev. Neurosci.* **20**, 567–594 (1997).
- Halata, Z. Sensory innervation of the hairy skin (light- and electronmicroscopic study). *J. Invest. Derm.* **101**, 75s–81s (1993).
- Waldmann, R. & Lazdunski, M. H⁺-gated cation channels: neuronal acid sensors in the NaC/DEG family of ion channels. *Curr. Opin. Neurobiol.* **8**, 418–424 (1998).
- Basiliana, F. *et al.* The acid-sensitive ionic channel subunit ASIC and the mammalian degenerin MDEG form a heteromultimeric H⁺-gated Na⁺ channel with novel properties. *J. Biol. Chem.* **272**, 28819–28822 (1997).
- Mano, I. & Driscoll, M. DEG/ENaC channels: a touchy superfamily that watches its salt. *Bioessays* **21**, 568–578 (1999).
- Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
- Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
- Lingueglia, E. *et al.* A modulatory subunit of acid sensing ion channels in brain and dorsal root ganglion cells. *J. Biol. Chem.* **272**, 29778–29783 (1997).
- Koltzenburg, M., Stucky, C. L. & Lewin, G. R. Receptive properties of mouse sensory neurons innervating hairy skin. *J. Neurophysiol.* **78**, 1841–1850 (1997).
- Carroll, P., Lewin, G. R., Koltzenburg, M., Toyka, K. V. & Thoenen, H. A role for BDNF in mechanosensation. *Nature Neurosci.* **1**, 42–46 (1998).
- French, A. S. Mechanotransduction. *Annu. Rev. Physiol.* **54**, 135–152 (1992).
- Koerber, H. R., Druzynski, R. E. & Mendell, L. M. Properties of somata of spinal dorsal root ganglion cells differ according to peripheral receptor innervation. *J. Neurophysiol.* **60**, 1584–1596 (1988).
- Krishtal, O. A. & Pidoplichko, V. I. A receptor for protons in the membrane of sensory neurons may participate in nociception. *Neuroscience* **6**, 2599–2601 (1981).
- Bevan, S. & Yeats, J. protons activate a cation conductance in a sub-population of rat dorsal root ganglion neurons. *J. Physiol. (Lond.)* **433**, 145–161 (1991).
- Steen, K. H., Reeh, P. W., Anton, E. & Handwerker, H. O. Protons selectively induce lasting excitation and sensitization to mechanical stimulation of nociceptors in rat skin, *in vitro*. *J. Neurosci.* **12**, 86–95 (1992).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel.

Science **287**, 2229–2234 (2000).

- Rusch, A. & Hummler, E. Mechano-electrical transduction in mice lacking the alpha-subunit of the epithelial sodium channel. *Hear. Res.* **131**, 170–176 (1999).
- Fricke, B. *et al.* Epithelial Na⁺ channels and stomatin are expressed in rat trigeminal mechanosensory neurons. *Cell Tissue Res.* **299**, 327–334 (2000).
- Drummond, H. A., Abboud, F. M. & Welsh, M. J. Localization of β and γ subunits of ENaC in sensory nerve endings in the rat foot pad. *Brain Res.* (in the press).
- McDonald, F. M. *et al.* Disruption of the β subunit of the epithelial Na⁺ channel in mice: hyperkalemia and neonatal death associated with a pseudohypoaldosteronism phenotype. *Proc. Natl Acad. Sci. USA* **96**, 1727–1731 (1999).
- Mannsfeldt, A. G., Carroll, P., Stucky, C. L. & Lewin, G. R. Stomatin, a MEC-2 like protein, is expressed by mammalian sensory neurons. *Mol. Cell. Neurosci.* **13**, 391–404 (1999).
- Benson, C. J., Eckert, S. P. & McCleskey, E. W. Acid-evoked currents in cardiac sensory neurons: a possible mediator of myocardial ischemic sensation. *Circ. Res.* **84**, 921–928 (1999).
- Stucky, C. L. & Lewin, G. R. Isolectin B₄-positive and -negative nociceptors are functionally distinct. *J. Neurosci.* **19**, 6497–6505 (1999).

Acknowledgements

We thank D. Melssen, E. Tarr, T. Moninger, R. Hrsta, T. Nesselhauf, P. Weber, A. Kanehl and T. Mayhew for assistance. We also thank the University of Iowa DNA Core Facility and the Central Microscopy Facility for assistance. This work was supported by the HHMI (M.J.W.) and a DFG grant (G.R.L.). P.A.H. was supported by a Marie Curie fellowship from the European Union. M.J.W. is an Investigator of the HHMI.

Correspondence and requests for materials should be addressed to M.J.W. (e-mail: mjwelsh@blue.weeg.uiowa.edu) or G.R.L. (e-mail: glewin@mdc-berlin.de).

Urinary bladder hyporeflexia and reduced pain-related behaviour in P2X₃-deficient mice

Debra A. Cockayne*, Sara G. Hamilton†, Quan-Ming Zhu*, Philip M. Dunn‡, Yu Zhong‡, Sanja Novakovic*, Annika B. Malmberg*, Gary Cain*, Amy Berson*, Laura Kassotakis*, Linda Hedley*, Wilhelm G. Lachnit*, Geoffrey Burnstock‡, Stephen B. McMahon† & Anthony P. D. W. Ford*

* The Neurobiology Unit, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, California 94304, USA

† Centre for Neuroscience Research, Kings College London, London SE1 9RT, UK

‡ Autonomic Neuroscience Institute, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, UK

Extracellular ATP is implicated in numerous sensory processes ranging from the response to pain to the regulation of motility in visceral organs¹. The ATP receptor P2X₃ is selectively expressed on small diameter sensory neurons^{2–4}, supporting this hypothesis. Here we show that mice deficient in P2X₃ lose the rapidly desensitizing ATP-induced currents in dorsal root ganglion neurons. P2X₃ deficiency also causes a reduction in the sustained ATP-induced currents in nodose ganglion neurons. P2X₃-null mice have reduced pain-related behaviour in response to injection of ATP and formalin. Significantly, P2X₃-null mice exhibit a marked urinary bladder hyporeflexia, characterized by decreased voiding frequency and increased bladder capacity, but normal bladder pressures. Immunohistochemical studies localize P2X₃ to nerve fibres innervating the urinary bladder of wild-type mice, and show that loss of P2X₃ does not alter sensory neuron innervation density. Thus, P2X₃ is critical for peripheral pain responses and afferent pathways controlling urinary bladder volume reflexes. Antagonists to P2X₃ may therefore have therapeutic potential in the treatment of disorders of urine storage and voiding such as overactive bladder.

The ion-channel subunit P2X₃ is one of seven known subunits that form homomeric and heteromeric receptors for ATP⁵. Uniquely, P2X₃ is expressed by a subgroup of small sensory neurons of the dorsal root and cranial ganglia^{2–4}. ATP and α,β-Me-ATP (a

P2X_{1,3}-selective analogue) excite many small calibre afferent neurons from skin, joints and viscera⁶. Studies on ATP release^{6,7} and P2 receptor antagonism^{8,9} support the involvement of P2X receptors in nociception and the micturition reflex. To investigate the role of P2X₃ in sensory processing, we generated P2X₃-deficient mice carrying a 1-kb deletion of the P2X₃ gene encompassing exon 1 and the initiating codon ATG (Fig. 1a, b).

Normally, P2X₃ is selectively expressed by small sensory neurons marked by the lectin IB4 (refs 4, 10). In P2X₃^{-/-} mice, P2X₃ immunoreactivity is undetectable in dorsal root ganglion (DRG), spinal cord and peripheral tissues (Fig. 1d, f, h, j). However, staining for IB4 in DRG and spinal cord (Fig. 1e–h), and protein gene product (PGP) 9.5 pan-neuronal staining in the skin (Fig. 1i, j) appears to be unaltered in P2X₃^{-/-} mice. Thus, loss of P2X₃ does not appear to affect either peripheral or central innervation patterns. Immunostaining for P2X receptor subunits P2X₂, P2X₅ and P2X₆ was also qualitatively unchanged in DRG, trigeminal and nodose ganglia of P2X₃^{-/-} mice (P2X₂ in DRG, Fig. 1c, d; and data not shown), indicating no compensatory changes in other P2X receptor subunits. P2X₁, P2X₄ and P2X₇ receptors were not detected in mouse sensory ganglia.

We analysed dissociated DRG and nodose ganglion neurons using

electrophysiology (Fig. 2a–d). In P2X₃^{+/+} DRG, 75% (27/38) and 60% (18/30) of neurons tested responded to ATP and α,β-Me-ATP, respectively. ATP-responsive DRG neurons showed either a rapidly desensitizing inward current (18/27, averaging 0.4 ± 0.08 nA, Fig. 2a), or a slowly desensitizing response (9/27, averaging 0.4 ± 0.22 nA, data not shown). No P2X₃^{-/-} DRG neurons (0/34) responded to either ATP or α,β-Me-ATP with rapidly desensitizing currents (Fig. 2b). The proportion of DRG neurons with slowly desensitizing responses to ATP (4/34, 12%) was not significantly different (*P* > 0.05) from that observed in wild-type mice, nor was the proportion of cells responding to GABA (γ-aminobutyric acid) or capsaicin (data not shown). Thus, P2X₃ homomers appear to be principally responsible for rapidly desensitizing ATP-activated currents in DRG. However, P2X₂ homomers and P2X_{2/3} heteromers may function in a minority of DRG neurons^{11–15}. In contrast, in P2X₃ wild-type mice, >90% (34/37) of nodose ganglion neurons responded to ATP and α,β-Me-ATP with slowly desensitizing persistent responses (Fig. 2c). In null-mutant mice, the proportion of nodose neurons responding to ATP (43/51, 84%) was similar (*P* > 0.1). However, the mean amplitude of the response (3.2 ± 0.6 nA for 100 μM ATP, *n* = 26) was significantly less (*P* < 0.05) than that observed for P2X₃^{+/+} neurons

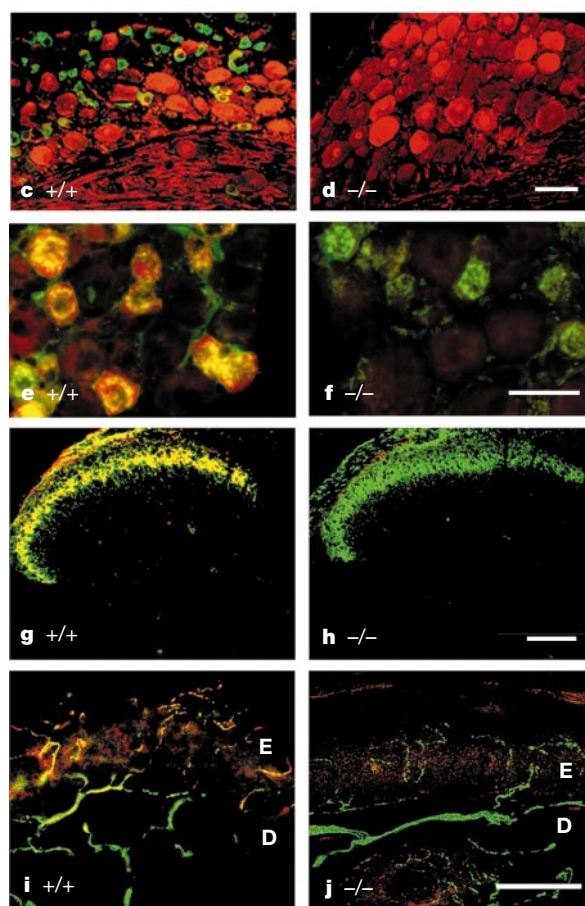
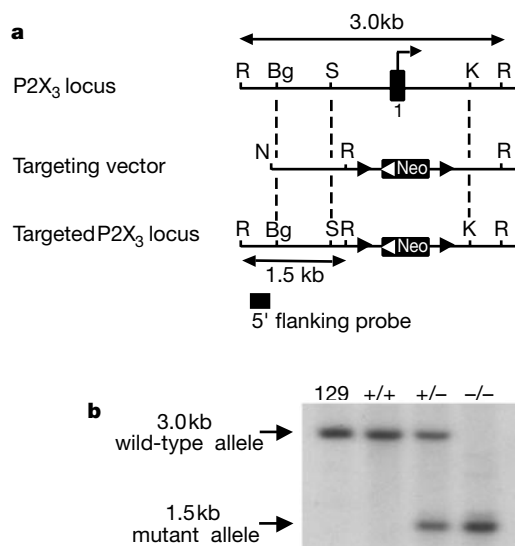


Figure 1 Targeted disruption of the P2X₃ gene and immunolocalization studies. **a**, Gene targeting strategy. R, *EcoR*I; Bg, *Bgl*I; S, *Sac*I; K, *Kpn*I; X, *Xho*I; N, *Not*I; Tk, thymidine kinase. **b**, Southern blot of P2X₃^{+/+}, P2X₃^{+/-} and P2X₃^{-/-} mice using a 5' flanking region probe shown in **a**. **c–j**, Colocalization of P2X₃ with neuronal markers in DRG, spinal cord and skin of P2X₃^{+/+} and P2X₃^{-/-} mice. **c, d**, Transverse sections (10 μm) of L5 DRG immunostained for P2X₃ (green) and P2X₂ (red). In P2X₃^{+/+} DRG, P2X₃ and P2X₂ immunoreactivities are present in small–medium and medium–large cells, respectively. P2X₂ staining appears unaltered in P2X₃^{-/-} DRG. **e–h**, Transverse sections of L5 DRG (15 μm; **e, f**) and lumbar spinal cord (20 μm; **g, h**) immunostained for P2X₃ (red) and IB4

lectin binding (green). In P2X₃^{+/+} DRG, nearly all P2X₃-immunoreactive cells bind IB4 (colocalization, yellow) and IB4 staining appears unaltered in P2X₃^{-/-} DRG. In P2X₃^{+/+} spinal cord, P2X₃ and IB4 staining terminals are co-localized (yellow) in inner lamina II of the dorsal horn, with apparently normal distribution of IB4 staining in P2X₃^{-/-} spinal cord. **i, j**, Sections (15 μm) of hindpaw plantar skin immunostained for P2X₃ (red) and the pan-neuronal marker PGP 9.5 (green). In P2X₃^{+/+} skin, P2X₃ and PGP 9.5 immunoreactivities are colocalized (yellow) in some fine epidermal (E) fibres, and to a lesser extent in nerve bundles in the dermis (D). Epidermal innervation is still evident by PGP 9.5 immunoreactivity in P2X₃^{-/-} mice. Scale bars: **c–f, i, j**, 25 μm; **g, h**, 75 μm.

(5.2 ± 0.5 nA for $100 \mu\text{M}$ ATP, $n = 31$) (Fig. 2d). None of the $\text{P2X}_3^{-/-}$ nodose neurons tested (0/12) responded to $\alpha, \beta\text{-Me-ATP}$ (Fig. 2d). Together with previous evidence^{3,16}, these data suggest that nodose ganglion neurons contain significant proportions of homomeric P2X_2 and heteromeric $\text{P2X}_{2/3}$ channels.

We next examined sensory deficits in P2X_3 wild-type and null-mutant mice. No differences were observed in locomotor activity and rotarod performance (data not shown). Injection of ATP into the hindpaw evoked a nociceptive behavioural response (intermittent hindpaw lifting, licking and biting, as in the rat¹⁷) in P2X_3 wild-type mice that was dose dependent (Fig. 2e). In P2X_3 null-mutant mice (Fig. 2e), responses were significantly decreased by 77% and 45% to 100 and 500 nmol of ATP, respectively. Altered pain responses were specific to ATP and not seen with intraplantar injections of $30 \mu\text{g}$ capsaicin (data not shown). Moreover, the non-selective P2 receptor antagonist pyridoxalphosphate-6-azophenyl-2',4'-disulphonic acid (PPADS) further reduced the residual hindpaw lifting behaviour of $\text{P2X}_3^{-/-}$ mice by about 50% (Fig. 2e). These data suggest that the pain-producing effects of peripheral ATP in humans^{18,19} and animals^{17,20} are mainly mediated by P2X_3 subunits. However, some of this response appears to be derived from other P2 receptors. Responses to noxious thermal and mechanical stimuli were similar in P2X_3 wild-type and null-mutant mice (data not shown). In contrast, pain-related behaviour was significantly attenuated in both phases of the formalin test ($\sim 50\%$) in null-mutant mice (Fig. 2f), consistent with previous work⁸.

We also investigated the role of P2X_3 in urinary bladder sensory function. Bladder afferent activity during filling drives micturition contractions mediated by the central nervous system (CNS)^{21,22}.

Therefore, we monitored these reflex contractions in P2X_3 wild-type and null-mutant mice using two different urodynamic methods. Figure 3a shows representative cystometrograms from conscious mouse cystometry studies in which voiding reflexes were measured in response to a continuous intravesical infusion of saline. $\text{P2X}_3^{-/-}$ null mice had significantly decreased micturition frequencies (mean void intervals 9.0 ± 0.8 min versus 5.3 ± 0.3 min in $\text{P2X}_3^{+/+}$ mice) (Fig. 3b, left), and significantly increased bladder capacities (mean void volumes 0.41 ± 0.04 ml versus 0.23 ± 0.02 ml in $\text{P2X}_3^{+/+}$ mice) (Fig. 3b, middle), but showed no differences in bladder pressures recorded at baseline, micturition threshold (not shown) and micturition peak (Fig. 3b, right). Male and female mice showed similar urodynamic changes. Additionally, cystometric differences are not attributable to influence from the 129Sv background (Fig. 3b).

Acute cystometry carried out under anaesthesia also showed micturition hyporeflexia in P2X_3 -null mice. Bladder contractions measured in response to distension with a fixed infusion rate of saline (Fig. 3c) resulted in frequent micturition contractions in $\text{P2X}_3^{+/+}$ mice, but virtually no contractions in $\text{P2X}_3^{-/-}$ mice, up to the cut-off volume. The average number of contractions per cystometrogram was significantly reduced in $\text{P2X}_3^{-/-}$ mice (0.6 ± 0.38 compared with 4.9 ± 2.37 in $\text{P2X}_3^{+/+}$ mice) (Fig. 3d). Accordingly, the micturition threshold was significantly increased (0.29 ± 0.01 compared with 0.21 ± 0.02 ml in $\text{P2X}_3^{+/+}$, $P < 0.05$), and only 23% of $\text{P2X}_3^{-/-}$ mice reached the micturition threshold compared with 75% of $\text{P2X}_3^{+/+}$ mice. No differences were observed in the intravesical pressure at a volume of 0.3 ml (data not shown).

Finally, we detected P2X_3 immunoreactivity on sensory neurons innervating the suburothelial nerve plexus of wild-type mouse

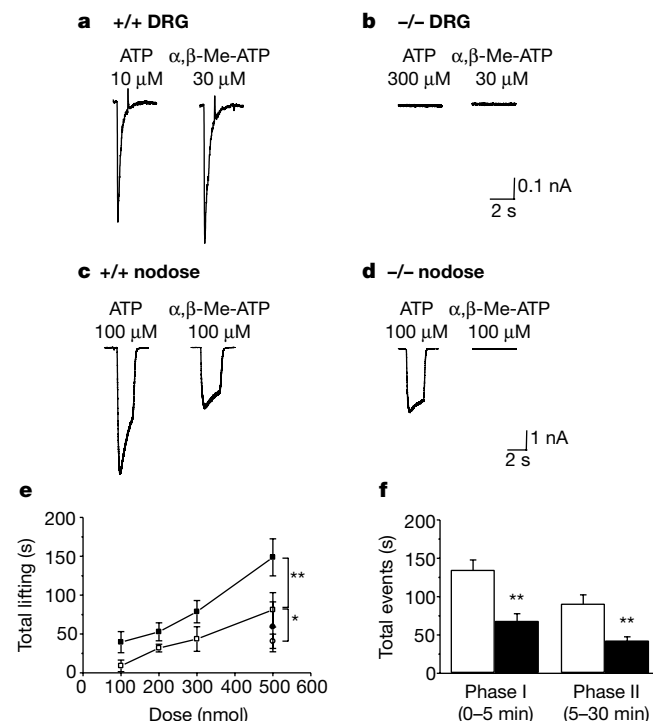


Figure 2 Responses to nucleotide agonists and nociceptive behaviour in P2X_3 -deficient mice. **a–d**, Whole-cell patch-clamp recordings of DRG and nodose ganglion neurons at a holding potential of -60 mV. **a**, $\text{P2X}_3^{+/+}$ DRG neurons show rapidly desensitizing inward currents in response to $10 \mu\text{M}$ ATP and $30 \mu\text{M}$ $\alpha, \beta\text{-Me-ATP}$. **b**, $\text{P2X}_3^{-/-}$ DRG neurons failed to produce a transient response to either $300 \mu\text{M}$ ATP or $30 \mu\text{M}$ $\alpha, \beta\text{-Me-ATP}$. **c**, $\text{P2X}_3^{+/+}$ nodose ganglion neurons show slowly desensitizing inward currents in response to $100 \mu\text{M}$ ATP and $\alpha, \beta\text{-Me-ATP}$. **d**, $\text{P2X}_3^{-/-}$ nodose ganglion neurons responded to $100 \mu\text{M}$ ATP, but not to $100 \mu\text{M}$ $\alpha, \beta\text{-Me-ATP}$. **e**, ATP-evoked behavioural responses. Hindpaw-lifting responses in 2–3-month-old male $\text{P2X}_3^{+/+}$ (filled square) and $\text{P2X}_3^{-/-}$ (open square) mice ($n = 10$ to 15) following intraplantar injection of varying doses

of ATP in a total volume of $20 \mu\text{l}$. Double asterisk, $P < 0.01$ for $\text{P2X}_3^{+/+}$ and $\text{P2X}_3^{-/-}$ mice; analysis of variance. The response to 500 nmol ATP was also measured 20 min following pre-treatment with a 200 mg per kg body weight intraperitoneal injection of PPADS in $\text{P2X}_3^{+/+}$ (filled circle) and $\text{P2X}_3^{-/-}$ (open circle) mice ($n = 10$ to 15). Asterisk, $P < 0.05$ for $\text{P2X}_3^{+/+}$ and $\text{P2X}_3^{-/-}$ mice; Student's t -test (significance shown only for $\text{P2X}_3^{-/-}$). **f**, Formalin-induced behavioural responses. Hindpaw-lifting and licking responses in 3–4-month-old male $\text{P2X}_3^{+/+}$ (open bars) and $\text{P2X}_3^{-/-}$ (filled bars) mice ($n = 10$) following intraplantar injection of 5% formalin in a total volume of $20 \mu\text{l}$. Double asterisk, $P < 0.01$ for $\text{P2X}_3^{+/+}$ and $\text{P2X}_3^{-/-}$ mice; Student's t -test.

bladder (Fig. 4c), both on small nerve fibres with terminals embedded in the urothelium, as well as on nerve bundles. Bladders from $P2X_3$ -null mice showed no $P2X_3$ immunoreactivity (Fig. 4d), but were otherwise histologically normal (Fig. 4a, b), and showed no alterations in sensory innervation patterns as measured by

capsaicin receptor (VR-1) immunoreactivity (Fig. 4e, f).

Our findings demonstrate the importance of $P2X_3$ receptors in somatic and visceral sensory function. First, we show that much of the DRG response to ATP is mediated by homomeric $P2X_3$ receptors, while in nodose ganglion neurons homomeric $P2X_2$ and

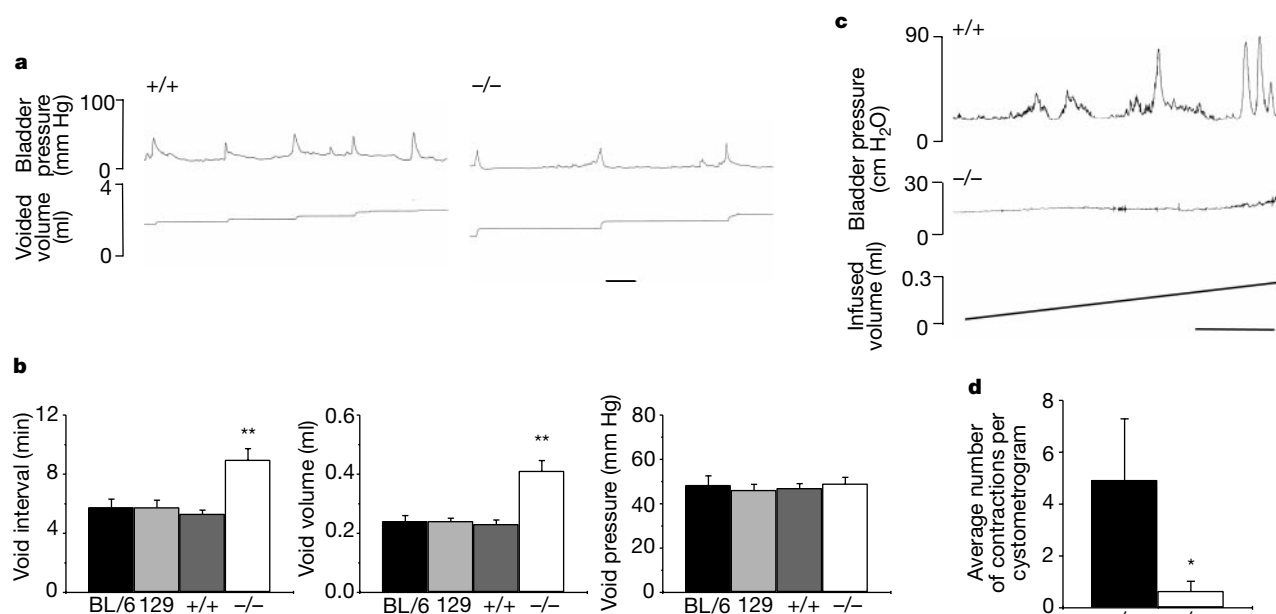


Figure 3 Bladder cystometry in $P2X_3$ -deficient mice. **a**, Representative cystometrograms from conscious 5–6-month-old $P2X_3^{+/+}$ and $P2X_3^{-/-}$ mice. Traces illustrate bladder pressure recorded in response to a constant intravesical infusion of saline ($50 \mu\text{l min}^{-1}$), and accumulated void volumes recorded from each micturition (scale bar, 2 min). **b**, Void intervals, void volumes and void pressures were quantified for C57BL/6 ($n = 7$), 129Sv ($n = 6$), $P2X_3^{+/+}$ ($n = 4$ males and 4 females) and $P2X_3^{-/-}$ ($n = 4$ males and 4 females) mice. $P2X_3^{-/-}$ mice had significantly decreased micturition frequencies (increased void interval) and significantly increased bladder capacities (increased void volume), but no

differences in bladder pressures. Double asterisk, $P < 0.01$ for $P2X_3^{+/+}$ and $P2X_3^{-/-}$ mice; analysis of variance. **c**, Representative acute cystometrograms recorded from anaesthetized and transurethral catheterized 5–6-month-old $P2X_3^{+/+}$ and $P2X_3^{-/-}$ mice. Each cystometrogram consisted of intravesical infusion of saline ($20 \mu\text{l min}^{-1}$ for 15 min) (scale bar, 2 min). Contractions greater than 10 cm H_2O were taken as micturition contractions. **d**, Quantification of the average number of contractions per cystometrogram for $P2X_3^{+/+}$ ($n = 8$) and $P2X_3^{-/-}$ ($n = 11$) mice confirms the altered micturition reflex in $P2X_3^{-/-}$ mice. Asterisk, $P < 0.05$; Student's t -test.

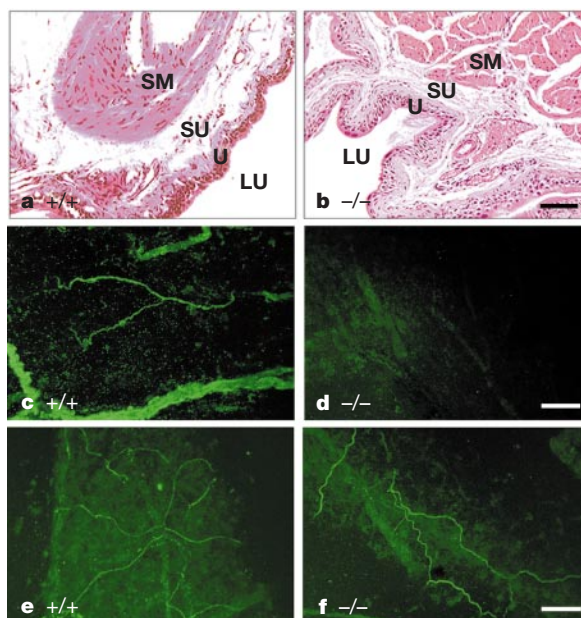


Figure 4 Immunolocalization in the mouse urinary bladder. **a, b**, Haematoxylin and eosin stained transverse sections ($10 \mu\text{m}$) showing different layers of the urinary bladder (near trigone): LU, lumen; U, urothelium; SU, suburothelium; SM, smooth muscle. Loss of $P2X_3$ does not cause degenerative or hyperplastic changes in $P2X_3^{-/-}$ bladder. **c, d**, Confocal images of whole mount bladder exposing the suburothelial sensory nerve plexus. In

$P2X_3^{+/+}$ bladder $P2X_3$ immunoreactivity is detected on small nerve fibres with terminals embedded in the urothelium, and on a large nerve bundle (**c**). In $P2X_3^{-/-}$ bladder, $P2X_3$ immunoreactivity is absent (**d**), but sensory innervation to the urinary bladder appears to be intact as evidenced by staining for the sensory neuron-specific capsaicin (VR-1) receptor (**e, f**). Scale bars, $50 \mu\text{m}$.

heteromeric P2X_{2/3} receptors appear most important. Second, our formalin test data are consistent with a role for ATP activation of P2X₃ in mediating some nociceptive responses to tissue damage. Finally, we show that P2X₃ is critical in regulating micturition reflex excitability. One explanation for these data is that ATP, released in response to stretch during distension and filling of the urinary bladder, excites primary afferent voiding circuitry through direct interaction with P2X₃ receptors. ATP is released from rabbit urothelium in response to stretch⁷, and P2X₃ is clearly present on nerve fibres innervating urinary bladder²³ (Fig. 4). Electrophysiological evidence also indicates that α,β -Me-ATP directly activates and desensitizes mechanosensitive pelvic afferents arising from rat urinary bladder⁹. Thus, loss of P2X₃ might impair sensory neuron activity during bladder filling, raising the volume threshold for activation of the micturition reflex. As loss of compliance and lowered volume thresholds are a component of many bladder storage disorders (for example, overactive bladder)²⁴, selective modulation of P2X₃ may provide new therapies. The potential for similar P2X₃ roles in mechanosensation in other hollow organs (for example, GI tract and lung)²⁵ needs to be explored. □

Methods

Physiological studies

F₂ and F₃ mice were used for *in vitro* and *in vivo* studies, respectively. All experiments were performed blind. Dissociation of neurons and whole-cell patch-clamp recording was carried out as described previously²⁶. Agonists were applied rapidly by microperfusion from a 4-barrel manifold controlled by computer-driven solenoid valves. Exchange of solution around the cell was complete in less than 100 ms. Time between applications was 2 min (nodose) and 3.5 min (DRG), allowing sufficient time to achieve reproducible responses. The minimum detectable response was 20 pA. Traces were acquired using FETCHEX (pCLAMP V.6.04 software, Axon Instruments), and plotted using ORIGIN V.4.1 (Microcal). Pain-related responses to injection of ATP into the hindpaw were measured essentially as described for rat¹⁷. The hindpaw lifting time was measured for a total of 4 min following injection of ATP. Thermal sensitivity was assessed using a radiant heat stimulus and tail immersion in a 52 °C water bath. Mechanical sensitivity was assessed using a set of calibrated von Frey filaments. For the formalin test, the hindpaw lifting and licking time was measured for a total of 30 min. Conscious mouse cystometry was performed essentially as described for rat²⁷. Recovery following catheter implantation was for 7 days, and intravesical saline infusion was at a rate of 50 μ l min⁻¹. For transurethral cystometry, bladder reflexes were assessed in urethane-anesthetized mice essentially as described for rat²⁸. Each cystometrogram consisted of intravesical distension to a total volume of 0.3 ml, at a rate of 20 μ l min⁻¹. Contractions greater than 10 cm of H₂O were taken as micturition contractions.

Generation of P2X₃ receptor-deficient mice and immunohistochemistry methods are described in Supplementary Information.

Received 19 May; accepted 17 August 2000.

1. Burnstock, G. P2X receptors in sensory neurons. *Br. J. Anaesth.* **84**, 476–488 (2000).
2. Chen, C. C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377**, 428–431 (1995).
3. Lewis, C. *et al.* Coexpression of P2X₂ and P2X₃ receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377**, 432–435 (1995).
4. Bradbury, E. J., Burnstock, G. & McMahon, S. B. The expression of P2X₃ purinoceptors in sensory neurons: Effects of axotomy and glial-derived neurotrophic factor. *Mol. Cell. Neurosci.* **12**, 256–268 (1998).
5. Ralevic, V. & Burnstock, G. Receptors for purines and pyrimidines. *Pharmacol. Rev.* **50**, 413–492 (1998).
6. Burnstock, G., McMahon, S. B., Humphrey, P. P. A. & Hamilton, S. G. In *Proceedings of the 9th World Congress on Pain* (eds Devor, M., Rowbotham, M. & Wiesenfeld-Hallin, Z.) 63–76 (IASP Press, Seattle, 2000).
7. Ferguson, D. R., Kennedy, I. & Burton, T. J. ATP is released from rabbit urinary bladder epithelial cells by hydrostatic pressure changes—a possible sensory mechanism? *J. Physiol. (Lond.)* **505**, 503–511 (1999).
8. Tsuda, M., Ueno, S. & Inoue, K. Evidence for the involvement of spinal endogenous ATP and P2X receptors in nociceptive responses caused by formalin and capsaicin in mice. *Br. J. Pharmacol.* **128**, 1497–1504 (1999).
9. Namiasivayam, S., Eardley, I. & Morrison, J. F. B. Purinergic sensory neurotransmission in the urinary bladder: an *in vitro* study in the rat. *Br. J. Urol. Int.* **84**, 854–860 (1999).
10. Vulchanova, L. *et al.* P2X₃ is expressed by DRG neurons that terminate in inner lamina II. *Eur. J. Neurosci.* **10**, 3470–3478 (1998).
11. Robertson, S. J., Rae, M. G., Rowan, E. G. & Kennedy, C. Characterization of a P2X₃-purinoceptor in cultured neurones of the rat dorsal root ganglia. *Br. J. Pharmacol.* **118**, 951–956 (1996).
12. Rae, M. G., Rowan, E. G. & Kennedy, C. Pharmacological properties of P2X₃-receptors present in neurones of the rat dorsal root ganglia. *Br. J. Pharmacol.* **124**, 176–180 (1998).
13. Burgard, E. C. *et al.* P2X receptor-mediated ionic currents in dorsal root ganglion neurons. *J. Neurophysiol.* **82**, 1590–1598 (1999).
14. Ueno, S., Tsuda, M., Iwanaga, T. & Inoue, K. Cell type-specific ATP-activated responses in rat dorsal root ganglion neurons. *Br. J. Pharmacol.* **126**, 429–436 (1999).

15. Grubb, B. D. & Evans, R. J. Characterization of cultured dorsal root ganglion neuron P2X receptors. *Eur. J. Neurosci.* **11**, 149–154 (1999).
16. Thomas, S., Virginio, C., North, R. A. & Surprenant, A. The antagonist trinitrophenyl-ATP reveals co-existence of distinct P2X receptor channels in rat nodose neurones. *J. Physiol. (Lond.)* **509**, 411–417 (1998).
17. Hamilton, S. G., Wade, A. & McMahon, S. B. The effects of inflammation and inflammatory mediators on nociceptive behaviour induced by ATP analogues in the rat. *Br. J. Pharmacol.* **126**, 326–332 (1999).
18. Blehan, T. & Keele, C. A. Observations on the algogenic actions of adenosine compounds on the human skin blister base preparation. *Pain* **3**, 367–377 (1977).
19. Hamilton, S. G., Warburton, J., Bhattacharjee, A., Ward, J. & McMahon, S. B. ATP in human skin elicits a dose related pain response which is potentiated under conditions of hyperalgesia. *Brain* **123**, 1238–1246 (2000).
20. Bland-Ward, P. A. & Humphrey, P. P. A. Acute nociception mediated by hindpaw P2X receptor activation in the rat. *Br. J. Pharmacol.* **122**, 365–371 (1997).
21. McMahon, S. B. In *Progress in Brain Research* Vol. 67 (eds Cervero, F. & Morrison, J. F. B.) 245–255 (Elsevier, Amsterdam, 1986).
22. de Groat, W. C. *et al.* Developmental and injury induced plasticity in the micturition reflex pathway. *Behav. Brain Res.* **92**, 127–140 (1998).
23. Elneil, S., Skepper, J. N., Kidd, E. J., Williamson, J. G. & Ferguson, D. R. The distribution of P2X₁ and P2X₃ receptors in the rat and human urinary bladder. *NeuroUrol. Urodyn.* **18**, 339–340 (1999).
24. Fitzgerald, J. M. & Krane, R. J. *The Bladder* (Churchill Livingstone; Edinburgh, London, Melbourne, New York and Tokyo; 1995).
25. Burnstock, G. Release of vasoactive substances from endothelial cells by shear stress and purinergic mechanosensory transduction. *J. Anat.* **194**, 335–342 (1999).
26. Zhong, Y., Dunn, P. M., Xiang, Z., Bo, X. & Burnstock, G. Pharmacological and molecular characterization of P2X receptors in rat pelvic ganglion neurons. *Br. J. Pharmacol.* **125**, 771–781 (1998).
27. Ishizuka, O., Mattiasson, A. & Andersson, K.-E. Role of spinal and peripheral alpha2 adrenoceptors in micturition in normal conscious rats. *J. Urol.* **156**, 1853–1857 (1996).
28. Dmitrieva, N., Shelton, D., Rice, A. S. & McMahon, S. B. The role of nerve growth factor in a model of visceral inflammation. *Neuroscience* **78**, 449–459 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Muraski for microinjection and colony management; J. Thompson, S. Bingham and J. Sutton for behavioural tests; M. Bardini for immunohistochemistry; and K. Gregrow for necropsy and pathology.

Correspondence and requests for material should be addressed to D.A.C. (e-mail: debra.cockayne@roche.com).

Warm-coding deficits and aberrant inflammatory pain in mice lacking P2X₃ receptors

Veronika Souslova*, Paolo Cesare*, Yanning Ding*, Armen N. Akopian*, Louise Stanfa†, Rie Suzuki†, Katherine Carpenter†, Anthony Dickenson†, Susan Boyce‡, Ray Hill‡, Daniela Nebenius-Oosthuizen§, Andrew J.H. Smith§, Emma J. Kidd|| & John N. Wood*

* Department of Biology and † Department of Pharmacology, University College London, London WC1E 6BT, UK

‡ Merck Sharp and Dohme Research Labs, Terlings Park, Essex CM20 2QR, UK

§ Centre for Genome Research, Edinburgh University, Edinburgh EH9 3JQ, UK

|| Welsh School of Pharmacy, Cardiff University, Cardiff CF1 3XF, UK

ATP activates damage-sensing neurons (nociceptors) and can evoke a sensation of pain¹. The ATP receptor P2X₃ is selectively expressed by nociceptors^{2,3} and is one of seven ATP-gated, cation-selective ion channels^{4–6}. Here we demonstrate that ablation of the P2X₃ gene results in the loss of rapidly desensitizing ATP-gated cation currents in dorsal root ganglion neurons, and that the responses of nodose ganglion neurons to ATP show altered kinetics and pharmacology resulting from the loss of expression of P2X_{2/3} heteromultimers. Null mutants have normal sensorimotor function. Behavioural responses to noxious mechanical and thermal stimuli are also normal, although formalin-induced

pain behaviour is reduced. In contrast, deletion of the P2X₃ receptor causes enhanced thermal hyperalgesia in chronic inflammation. Notably, although dorsal-horn neuronal responses to mechanical and noxious heat application are normal, P2X₃-null mice are unable to code the intensity of non-noxious 'warming' stimuli.

To examine the physiological role of the P2X₃ receptor in the absence of selective antagonists, null-mutant mice were generated on an inbred (C57Bl6) and outbred (MF1) genetic background^{7,8}

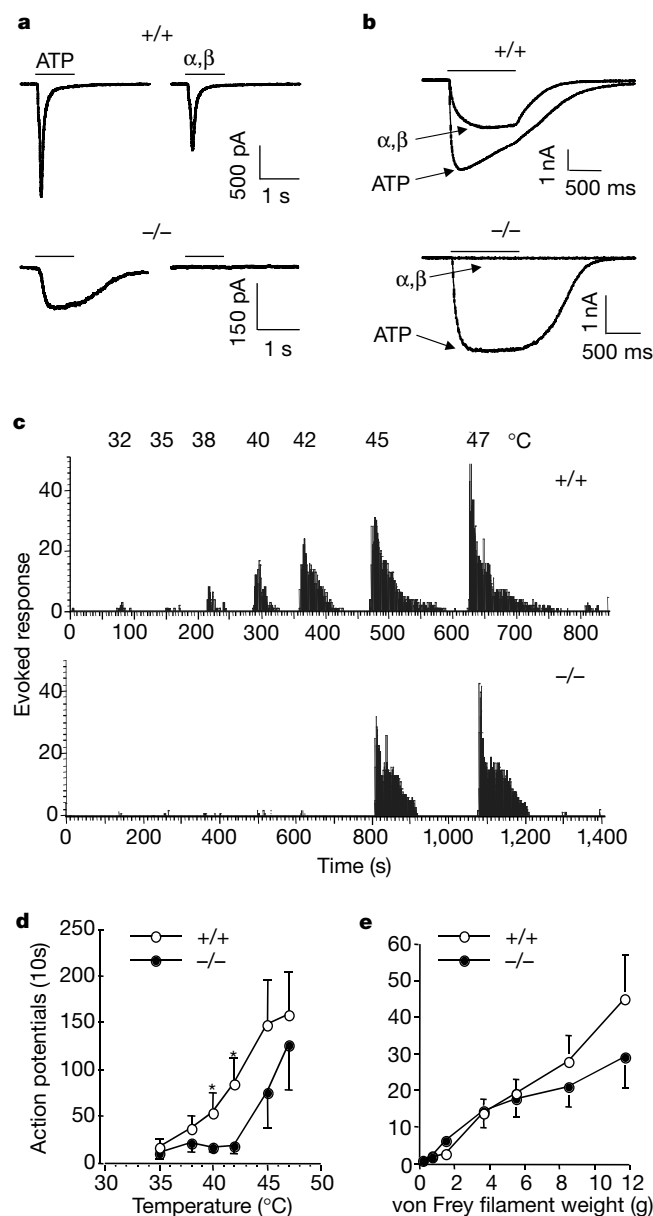


Figure 1 Physiological effects of deleting the P2X₃ receptor. **a, b**, The responses to a 1-s application of 10 μ M ATP in wild-type (+/+) and P2X₃-null (-/-) neurons. Cells responding to ATP were subsequently treated with 10 μ M α,β -Me-ATP (α,β). **a**, DRG neurons; **b**, nodose ganglion neurons. **c–e**, Responses of dorsal horn neurons to peripheral stimuli in wild-type and P2X₃-null mice. **c**, Individual examples of neuronal responses to graded thermal stimuli recorded in wild-type and P2X₃-null mice. **d**, Thermal stimulus-response curves recorded in wild-type ($n = 12$) and P2X₃-null ($n = 17$) mice. Responses are presented as mean number of action potentials $\pm$ s.e.m. Asterisk, $P \leq 0.05$. **e**, Responses to mechanical stimuli elicited with von Frey filaments show no significant differences between wild-type and P2X₃-null mice. Electrical thresholds and responses to suprathreshold electrical stimulation of A and C fibre primary afferents are shown in the Supplementary Information.

(see Supplementary Information). In most wild-type DRG neurons (87.5%, $n = 32$; Fig. 1a), ATP elicited a large inward membrane current (amplitude: 797 ± 122 pA, mean $\pm$ s.e.m., $n = 28$) that desensitized rapidly ($72 \pm 6\%$ desensitization in 1 s). In P2X₃-null neurons, ATP evoked an inward current in a smaller proportion of cells (56.7%, $n = 37$), the mean current was smaller (151 ± 52 pA; $n = 21$) and less desensitized ($11 \pm 2\%$ desensitized after 1 s). In wild-type DRG neurons, 88.5% ($n = 26$) of the cells activated by ATP responded to a subsequent application of α,β -Me-ATP with a large (464 ± 85 pA; $n = 23$) and rapidly desensitizing current ($87 \pm 4\%$). In P2X₃-null DRG, of the 20 neurons responding to ATP only 1

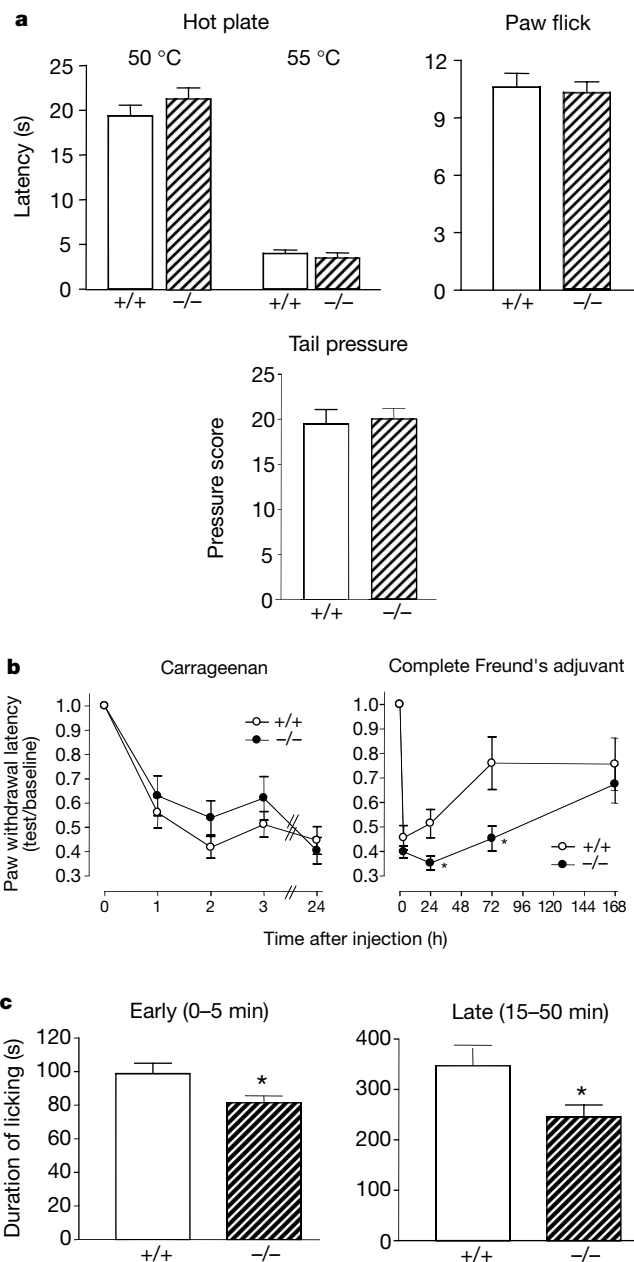


Figure 2 Behavioural effects of deleting the P2X₃ receptor. Sensorimotor function was unaffected (see Supplementary Information). Responses to acute noxious stimuli were normal (**a**, $n \geq 10$). Carrageenan-induced thermal hyperalgesia developed to similar levels in wild-type and P2X₃-null mice, but longer term thermal hyperalgesia induced by complete Freund's adjuvant was potentiated ($P < 0.05$, $n = 12$) in null mutants on two genetic backgrounds (**b**) and formalin-induced pain behaviour was attenuated in both phases (**c**, $P < 0.05$, $n = 10$). Asterisk, $P < 0.05$.

responded to α,β -Me-ATP, with a small (22 pA) and non-desensitizing current. This indicates that homomeric P2X₃ receptors are the major P2X subtype in mouse DRG neurons.

Patch-clamp recordings of nodose ganglion neurons in normal and P2X₃-null mice revealed very large ATP-evoked non-desensitizing currents. In nodose ganglion neurons (Fig. 1b), peak current had mean values of $3,365 \pm 428$ pA in wild-type ($n = 27$) and $2,250 \pm 684$ ($n = 23$) in P2X₃^{-/-} mice. In wild-type mice, desensitization was $25.9 \pm 4.4\%$ ($n = 27$), but P2X₃^{-/-} neurons desensitized more slowly ($7.6 \pm 4.7\%$; $n = 23$). In the wild type, α,β -Me-ATP application induced a mean current of $1,705 \pm 255$ pA ($n = 27$). In the null mutant, α,β -Me-ATP application failed to evoke any response ($n = 23$). Almost half of the currents evoked by ATP are due to P2X_{2/3} heteromultimers; P2X₂ receptors, as well as other P2X subtypes, must account for the rest of the current. The lack of P2X₁ receptors is confirmed by the fact that in P2X₃^{-/-} nodose ganglion neurons, α,β -Me-ATP was inactive.

In vivo recordings of dorsal horn neurons were used to examine the consequences of ablating P2X₃ on peripheral input into the spinal cord⁹. There was no significant change in the mechanically evoked responses in P2X₃-null mice, although the responses to the greatest pressure (8.5 and 11.7 g) were slightly reduced. Both threshold and suprathreshold responses of dorsal horn neurons evoked after electrical activation of A and C fibre primary afferents were unaltered in the P2X₃-null mice (see Supplementary Information). This indicates that centrally located P2X₃ receptors are not significantly involved in spinal nociceptive processing under acute conditions, consistent with the lack of effect of spinally administered P2X antagonists on nociceptive neuronal responses and behaviour measured in rats^{10–12}.

Differences were seen between the P2X₃-null and wild-type mice in their response to thermal stimuli, although no change was observed in the neuronal responses to mechanical stimuli (Fig. 1). Application of a range of thermal stimuli to the hindpaw of wild-type mice, evoked a graded neuronal response in the dorsal horn (Fig. 1c–d). In contrast, in P2X₃-null mice very little neuronal activity was evoked until noxious temperatures were reached¹³. This is in agreement with a 'normal' behavioural response to temperatures within the noxious range, consistent with behavioural studies showing no effect of pyridoxalphosphate-6-azophenyl-2',4'-disulphonic acid (PPADS) in acute tests of thermal nociception¹², but indicates a deficit in the ability of the null-mutant mice to code the intensity of non-noxious 'warming' stimuli.

Behavioural tests showed that responses to acute mechanical and thermal stimuli were normal, precluding a major role for P2X₃ in transducing acute noxious stimuli (Fig. 2, and see Supplementary Information). However formalin-induced pain behaviour was significantly reduced in both phases, consistent with data obtained using PPADS in mice¹⁴. Most strikingly, the development of thermal hyperalgesia in response to chronic inflammation induced with complete Freund's adjuvant but not short-term inflammatory stimuli such as carrageenan (Fig. 2b) or capsaicin (data not shown) was markedly potentiated.

These data indicate that P2X₃ receptors within the dorsal horn have a significant regulatory role in persistent inflammatory pain. In spinal cord slices, ATP has been shown to potentiate glutamate-induced and synaptically evoked currents, as well as causing delayed depression of synaptic currents, probably owing to the actions of adenosine^{11,12,15,16}. The unexpected hyperalgesia present in P2X₃ receptor knockouts may therefore be due to the actions of ATP on other less rapidly desensitizing P2 receptors within the dorsal horn, or may result from the loss of a regulatory function of P2X₃ receptors on other modulatory systems. Thus, there is no major role for the P2X₃ receptor in noxious mechanosensation and acute pain responses^{17,18}. However, these studies demonstrate a role for ATP acting through P2X₃ in inflammatory pain processing, as well as an unsuspected role for P2X₃ receptors in the coding of innocuous warmth. □

Methods

Gene targeting

A targeting construct in which exons 2–7 of the P2X₃ gene were deleted was constructed and knockout mice were generated and analysed by standard methods (see Supplementary Information)^{7,8,19}.

Electrophysiology

DRG or nodose ganglia from adult mice (6–9 weeks old) were removed and neurons were cultured²⁰. Recordings were made 1–3 days after plating, using patch-clamp electrodes of tip diameter 0.7 μ m and an Axopatch 200B amplifier (Axon Instruments) as described²⁰. Drugs were used at 10 μ M ATP (magnesium salt), 10 μ M α,β -Me-ATP (lithium salt). Cells between 15- and 35- μ m diameter were randomly chosen for experiments²⁰.

Dorsal horn recordings

Extracellular recordings were made *in vivo* as described^{10,21}. The neuronal responses evoked by transcutaneous electrical (train of 16 stimuli, 0.5 Hz, 2-ms wide pulse) and thermal (water jet applied for 10 s, 32–47 °C) stimuli applied to the peripheral receptive field located on the ipsilateral hind paw were characterized. Mechanical stimulation (von Frey filaments 0.166 g to 11.7 g) in the innocuous and noxious ranges was also applied to peripheral receptive fields.

Behavioural studies

Mice were examined for spinal reflexes and motor skills, as well as responses to acute and inflammatory pain, as described in detail²¹.

Received 19 May; accepted 17 August 2000.

- Bleehen T. & Keele, C. A. Observations on the algogenic actions of adenosine compounds on the human blister base preparation. *Pain* **3**, 367–377 (1977).
- Chen, C. *et al.* A P2X purinoceptor expressed by a subset of sensory neurons. *Nature* **377**, 428–431 (1995).
- Lewis, C. *et al.* Co-expression of P2X₂ and P2X₃ receptor subunits can account for ATP-gated currents in sensory neurons. *Nature* **377**, 432–435 (1995).
- MacKenzie, A. B., Surprenant, A. & North, R. A. Functional and molecular diversity of purinergic ion channel receptors. *Ann. NY Acad. Sci.* **30**, 716–729 (1999).
- Grubb, B. D. & Evans, R. J. Characterization of cultured dorsal root ganglion neuron P2X receptors. *Eur. J. Neurosci.* **11**, 149–154 (1999).
- Ueno, S., Tsuda, M., Iwanaga, T. & Inoue, K. Cell type-specific ATP-activated responses in rat dorsal root ganglion neurons. *Br. J. Pharmacol.* **126**, 429–436 (1999).
- Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
- Mogil, J. S. *et al.* Heritability of nociception I: responses of 11 inbred mouse strains on 12 measures of nociception. *Pain* **80**, 67–82 (1999).
- Chapman, V., Ng, J. & Dickenson, A. H. A novel spinal action of mexiletine in spinal somatosensory transmission of nerve injured rats. *Pain* **77**, 289–296 (1998).
- Stanfa, L. C., Kontinen, V. K. & Dickenson, A. H. Effects of spinally administered P2X receptor agonists and antagonists on the responses of dorsal horn neurones recorded in normal, carrageenan-inflamed and neuropathic rats. *Br. J. Pharmacol.* **129**, 351–359 (2000).
- Li, J. & Perl, E. R. ATP modulation of synaptic transmission in the spinal substantia gelatinosa. *J. Neurosci.* **15**, 3357–3365 (1995).
- Li, P., Calejesan, A. A. & Zhua, M. ATP P2X receptors and sensory synaptic transmission between primary afferent fibres and spinal dorsal horn neurons in rats. *J. Neurophysiol.* **80**, 3356–3360 (1998).
- Cesare, P. & McNaughton, P. Peripheral pain mechanisms. *Curr. Opin. Neurobiol.* **7**, 493–499 (1997).
- Tsuda, M., Ueno, S. & Inoue, K. Evidence for the involvement of spinal endogenous ATP and P2X receptors in nociceptive responses caused by formalin and capsaicin in mice. *Br. J. Pharmacol.* **128**, 1497–1504 (1999).
- Gu, J. G., Bardoni, R., Magherini, P. C. & MacDermott, A. B. Effects of the P2-purinoceptor antagonists suramin and PPADS on glutamatergic synaptic transmission in rat dorsal horn neurons of the spinal cord. *Neurosci. Lett.* **253**, 167–170 (1998).
- Gu, J. G. & MacDermott, A. B. Activation of ATP P2X receptors elicits glutamate release from sensory neuron synapses. *Nature* **389**, 749–753 (1997).
- Bland-Ward, P. A. & Humphrey, P. P. A. Acute nociception mediated by hindpaw P2X receptor activation in the rat. *Br. J. Pharmacol.* **122**, 365–371 (1997).
- Nakamura, F. & Strittmatter, S. M. P2Y₁ purinergic receptors in sensory neurons: contribution to touch-induced impulse generation. *Proc. Natl Acad. Sci. USA* **17**, 10465–10470 (1996).
- Kidd, E. J., Miller, K. J., Sansum, A. J. & Humphrey, P. P. Evidence for P2X₃ receptors in the developing rat brain. *Neuroscience* **87**, 533–539 (1998).
- Cesare, P., Souslova, V., Akopian, A., Rufan, O. & Wood, J. N. Fast inactivating responses to ATP are lost in dorsal root ganglion neurons isolated from P2X₃ knock-out mice. *Proc. Physiol. Soc.* **523P**, (2000).
- Akopian, A. N. *et al.* The tetrodotoxin-resistant sodium channel SNS plays a specialised role in pain pathways. *Nature Neurosci.* **2**, 541–548 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the Wellcome Trust, EC Biomed and the MRC for support and H. Thompson for help with genotyping mice.

Correspondence and requests for materials should be addressed to J.N.W. (e-mail: J.Wood@ucl.ac.uk).

Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*

Lúcio H. Freitas-Junior*, Emmanuel Bottius†, Lindsay A. Pirrit*, Kirk W. Deitsch‡, Christine Scheidig*, Françoise Guinet†, Ulf Nehrbass§, Thomas E. Wellems‡ & Artur Scherf*

* Unité de Biologie des Interactions Hôte-Parasite, CNRS URA 1960, and § Laboratoire de Biologie Cellulaire du Noyau, CNRS URA 1773, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

‡ Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA

Persistent and recurrent infections by *Plasmodium falciparum* malaria parasites result from the ability of the parasite to undergo antigenic variation and evade host immune attack^{1,2}. *P. falciparum* parasites generate high levels of variability in gene families that comprise virulence determinants of cytoadherence and antigenic variation^{3–7}, such as the *var* genes. These genes encode the major variable parasite protein (PfEMP-1), and are expressed in a mutually exclusive manner at the surface of the erythrocyte infected by *P. falciparum*^{8–12}. Here we identify a mechanism by which *var* gene sequences undergo recombination at frequencies much higher than those expected from homologous crossover events alone¹³. These recombination events occur between subtelomeric regions of heterologous chromosomes, which associate in clusters near the nuclear periphery in asexual blood-stage parasites or in bouquet-like configurations near one pole of the elongated nuclei in sexual parasite forms. We propose that the alignment of *var* genes in heterologous chromosomes facilitates gene conversion and promotes the diversity of antigenic and adhesive phenotypes. The association of virulence factors with a specific nuclear subcompartment may also have implications for variation during mitotic recombination in asexual blood stages.

The *var* genes of *P. falciparum* display great sequence diversity and are responsible for strain-specific differences in polymorphic erythrocyte surface antigens that affect pathogenesis and disease outcome^{7,14}. Individual parasites contain 50 or more *var* genes, most of which are located close to telomeric chromosome ends^{15–17}. In a genetic cross between two parasite clones (HB3×Dd2), non-parental *var* gene restriction fragments were detected much more frequently than expected from the estimated recombination frequency¹³ (Fig. 1a).

We investigated *var* inheritance further by isolating specific *var* gene tags from individual HB3 parasite chromosomes and mapping 18 of these tags to the chromosomes of *P. falciparum* clones HB3, Dd2 and 3D7. All 18 tags hybridized specifically to the chromosomes from which they were originally isolated except tag *HB3var12-1*, which hybridized to each of two chromosomes, 12 and 1 (see Table 1). Few of these tags cross-hybridized to chromosomes from 3D7 and Dd2 and two available *var* tags from the 3D7 strain hybridized exclusively to 3D7 chromosomes (Table 1). These results show that parasites from genetically diverse backgrounds can contain minimal overlapping *var* gene sets, as predicted from modelling studies for antigenic determinants of human pathogens¹⁸.

We chose 13 independent recombinant progeny from the HB3×

Dd2 cross and 8 progeny from the HB3×3D7 cross for initial analysis. As we expected, independent chromosome segregation had produced new *var* assortments in the progeny (Fig. 1b). Skewed inheritance was evident with some *var* genes. For example *HB3var2-1* was found in only 2 of the 13 progeny from the HB3×Dd2 cross and once in the 8 progeny from the HB3×3D7 cross, whereas *HB3var3/4-1* was found in 12 of the 13 progeny from the HB3×Dd2 cross (Table 1).

Our analysis of the position of the *var* gene tags in the chromosomes of the HB3×Dd2 progeny showed that, after crossing, *var* genes were not always found on the same chromosome as in the parental strain. The *HB3var10-1* gene tag, present only in chromosome 10 of the HB3 parental strain and certain progeny (such as B4R3 and SC01), hybridized in other progeny to both chromosomes

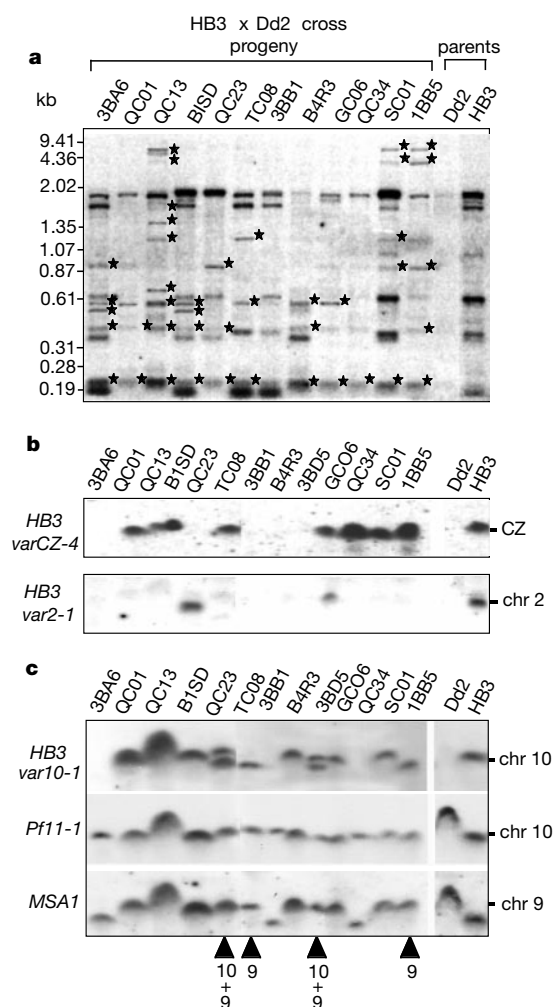


Figure 1 Frequent ectopic recombination events between members of the *var* gene family during sexual reproduction. **a**, Southern blot of *TaqI* digests of genomic DNA of parental strains HB3 and Dd2 and progeny clones. Almost all DBL- α sequences of a given *var* repertoire can be polymerase chain reaction (PCR) amplified in a single reaction using a set of universal primers. A mixture of these DBL- α sequences amplified from genomic DNA of the HB3 strain was used for the hybridization studies. Non-parental bands are marked with an asterisk. **b**, Signals from hybridization of *var* tags *HB3varCZ-4* (top) and *HB3var2-1* (bottom) to PFGE blots of progeny chromosomes from the HB3×Dd2 cross. The nomenclature of individual *var* genes includes the name of the strain and the chromosome location. CZ indicates the compression zone, a region that contains several co-migrating chromosomes. **c**, Signals from hybridization of the *HB3var10-1* tag to progeny chromosomes from the HB3×Dd2 crosses. Progeny clones with ectopic recombination events are indicated by arrowheads. The blot was stripped and re-hybridized with the chromosome-10-specific gene probe *Pf11-1* (middle) and the chromosome-9-specific probe *MSA1* (bottom).

† Present address: Bactériologie Moléculaire et Médicale, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France (E.G.); Gen Odyssee, Les Ulis 91974 Courtaboeuf Cedex, France (E.B.).

10 and 9 (such as QC23 and 3BD5), to chromosome 9 only (such as TC08 and 1BB5), or to no chromosome (such as QC34 and 3BB1) (Fig. 1c). Similar shifts were found in 5 out of 8 progeny clones from the HB3×3D7 cross (data not shown). These recombination events were mapped to the subtelomeric regions of chromosomes 9 and 10 by two-dimensional PFGE analysis (data not shown).

To investigate the molecular mechanism involved in the *HB3var10-1* non-allelic (ectopic) rearrangements, we performed Southern blot analysis of DNA from the parental HB3 and progeny clones. We identified a 3.3-kilobase (kb) band from chromosome 10 of the HB3 clone by hybridizing the *HB3var10-1* tag to DNA that was digested with the restriction enzyme *EcoRI*, which has sites outside the *var* DBL- α coding region. In the progeny carrying a duplication of the *HB3var10-1* sequence on chromosome 9, two hybridization bands were observed, the parental 3.3-kb band and a new fragment of 6 kb (progeny 3BD5, QC23, XP3 and X51; Fig. 2a). We observed only a single 6-kb band from progeny carrying the *HB3var10-1* gene in translocation from chromosome 10 to chromosome 9 (TC08, 1BB5, X12, X10 and XP2; Fig. 2a).

The genomic DNA of progeny from an HB3×HB3 self-cross (HB3-B) was also checked for recombination by *HB3var10-1* hybridization analysis. After high-stringency washes, the *HB3var10-1* probe detected the parental 3.3-kb fragment as well as the 6-kb fragment in HB3-B progeny parasites, whereas only the 3.3-kb band was detected in the original HB3 parent DNA (HB3-A, Fig. 2b). After low-stringency washes, the probe detected cross-hybridization with other sequences, suggesting at least one other *var* duplication event in HB3-B progeny (1.35-kb band, Fig. 2b). As straightforward allelic recombination would not be expected to generate new *var* forms in crosses between identical parents, these results provide additional evidence for ectopic gene conversion events in meiosis.

To explore mechanisms that may underlie the ectopic DNA rearrangements of telomere-associated *var* genes, we extended the DNA sequence of the *HB3var10-1* tag from parental HB3 chromosome 10 and progeny X10 clone chromosome 9. Specific primers based on the *HB3var10-1* tag sequence and vectorette linkers were used to amplify regions 3' of the known sequences. DNA sequence alignments showed identical sequences for 470 bp, with divergence further downstream (Fig. 2c). The beginning of this divergence

coincides with the 5' CIDR domain, corresponding to a region of the semi-conserved PfEMP1 head structure that can mediate binding to blood cells and the surface of endothelium.

Two potential mechanisms could be responsible for the observed hybridization patterns: gene conversion leading to duplication of the *HB3var10-1* fragment, or reciprocal exchange resulting in its movement from chromosome 10 to chromosome 9 (Fig. 2d). To evaluate these mechanisms, we analysed the original 13 and an additional 7 HB3×Dd2 progeny that have translocations or duplications of the *HB3var10-1* sequence. In these progeny, inheritance of the *HB3var10-1* sequence was compared to those of nearby dimorphic markers from the linkage groups of chromosomes 9 and 10 in the HB3×Dd2 genetic map¹³. Analysis of the chromosomal inheritance in progeny displaying the ectopic rearrangements showed that the *HB3var10-1* fragment was always duplicated

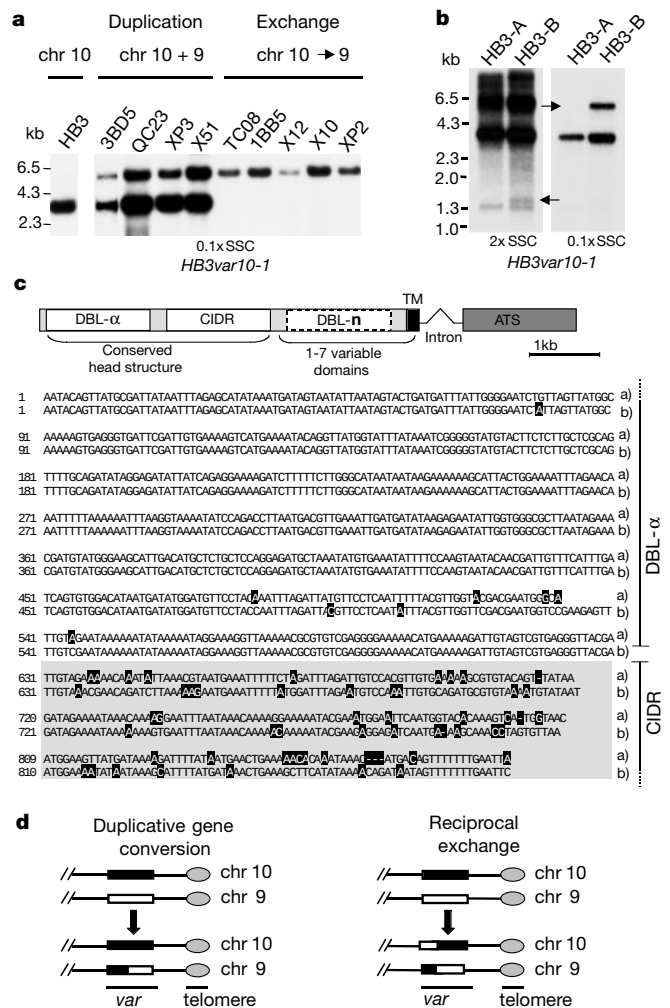


Figure 2 Ectopic gene conversion in progeny from the HB3×Dd2 and HB3×3D7 crosses. **a**, *EcoRI* fragments from independent recombinant progeny recognized by the *HB3var10-1* gene probe. **b**, Signals from *EcoRI*-digested DNA from the original HB3 parent (HB3-A) or progeny from the HB3 self-cross (HB3-B) after low and high stringency washes. *var* gene DNA rearrangements are indicated by arrows. **c**, DNA sequence alignment comparing the extended *HB3var10-1* sequence from the HB3 parent (**a**) against the inserted sequence (*X10var9-1*) isolated from chromosome 9 in the X10 progeny clone (**b**). Mismatches are shown shaded in black. The cysteine-rich interdomain region (CIDR) region is shown shaded in grey. A diagram of *var* genes is shown. **d**, Representations of possible mechanisms for ectopic *var* gene recombination. The favoured model of a gene conversion event would lead to a partial duplication of a *var* gene (left), whereas the model of reciprocal DNA recombination would result in two new *var* gene forms (right).

Table 1 Hybridization of *var* gene tags to laboratory strains and to independent progeny clones of laboratory crosses

<i>var</i> gene tag	Chromosome location			Inheritance in progeny	
	HB3	Dd2	3D7	HB3 × Dd2	HB3 × 3D7
<i>HB3var14-1</i>	14	No	No	n.d.	n.d.
<i>HB3var13-1</i>	13	No	No	n.d.	n.d.
<i>HB3var12-1</i>	12+	No	No	3/13	1/7
	1	No	No	5/13	5/7
<i>HB3var12-2</i>	12	No	4	5/12	3/8
					4/8
<i>HB3var12-3</i>	12	No	No	1/13	1/8
<i>HB3var12-4</i>	12	No	4	1/13	1/7
					3/7
<i>HB3var11-1</i>	11	No	No	n.d.	n.d.
<i>HB3var10-1</i>	10	No	No	10/13	8/8
<i>HB3varCZ-1</i>	5-9	No	No	6/13	1/8
<i>HB3varCZ-2</i>	5-9	No	No	2/13	1/7
<i>HB3varCZ-3</i>	5-9	No	No	6/13	5/8
<i>HB3varCZ-4</i>	5-9	No	No	8/13	5/8
<i>HB3var3/4-1</i>	3/4	No	No	12/13	4/7
<i>HB3var3/4-2</i>	3/4	No	No	7/13	n.d.
<i>HB3var3/4-3</i>	3/4+	No	No	5/13	3/7
<i>HB3var3/4-4</i>	3/4	No	No	4/13	n.d.
<i>HB3var2-1</i>	2	No	No	2/13	7/7
<i>HB3var1-1</i>	1	No	No	n.d.	n.d.
<i>3D7var12-1</i>	No	No	12	n.d.	7/8
<i>3D7var12-2</i>	No	No	12	n.d.	4/8

Results from hybridizations of *var* gene tags to chromosomes from the parents and progeny of the *P. falciparum* HB3×3D7 and HB3×Dd2 crosses. The *var* tag name indicates the strain from which the tag was isolated and the chromosome to which the tag hybridizes. The chromosomes to which the tags hybridize are indicated in the first column and the number of times each tag occurs in progeny from each cross is indicated in the second column. n.d., not determined.

from chromosome 10 into the same position within the chromosome 9 segment that was inherited from the HB3 parent.

If this duplication resulted from reciprocal exchange, progeny would have resulted that lacked the *HB3var10-1* sequence in the chromosome 10 segment from HB3. Such progeny were not observed. Instead, progeny that hybridized to chromosome 9 but not chromosome 10 (TC-08 and 1BB5) were found to have inherited chromosome 10 from the Dd2 parent. These data favour gene conversion between chromosomes 9 and 10 of HB3 as the primary mechanism responsible for the *HB3var10-1* rearrangements.

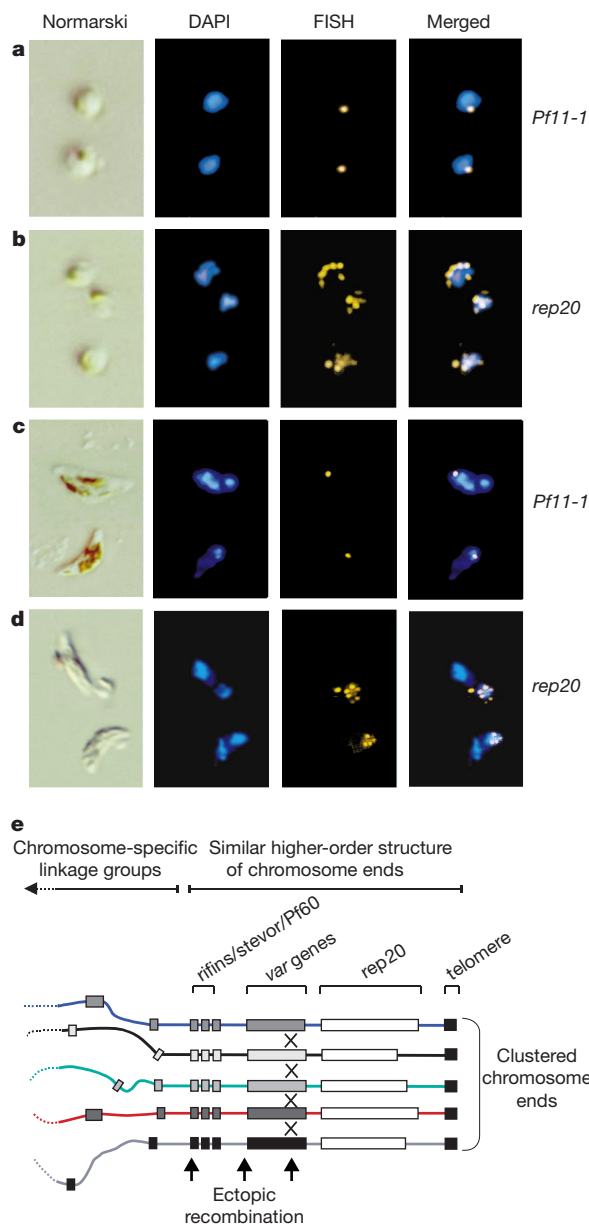


Figure 3 Physical clustering of *P. falciparum* subtelomeric regions detected in sexual and asexual parasites. Asexual (trophozoite) (**a**, **b**) and sexual blood-stage (gametocyte) (**c**, **d**) signals detected by FISH. The first column of panels shows Nomarski images of parasites, the second shows fluorescent images after DAPI staining (blue), and the third shows FISH signals (yellow) from the single copy *Pf11-1* gene probe or the *P. falciparum* telomere-associated *rep20* probe. The FISH and DAPI signals are superimposed in the fourth column. Nuclei of gametocytes present a distinct morphology, with chromosome ends forming bouquet-like structures. **e**, Representation of the organization of *P. falciparum* chromosome ends, including the telomere repeats, *rep20* and members of gene families such as the *var*, *rifins* and *STEVOR*^{16,17}.

Genes located at telomeric chromosome ends are often part of specialized higher order chromatin structures¹⁹. In budding yeast, telomeres have been shown to cluster in 'silent domains' at the periphery of nuclei, and in many species from yeast to mammals, sexual-stage and meiotic telomeres are found to be polarized forming a 'bouquet'^{20,21}. The alignment of telomeres supported by this structure has previously been reported to be important for recombination²². We therefore analysed the distribution of chromosome ends by fluorescence *in situ* hybridization (FISH) to determine whether *var* genes could be associated with such telomere clusters in *P. falciparum*. The telomere-associated, highly repetitive subtelomeric element *rep20* was used to probe asexual-stage parasites of 3D7, HB3 and FCR3 strains.

We observed between 4 and 7 spots of *rep20* hybridization in the nuclei of ring and trophozoite blood-stage parasites, indicating clustering of heterologous chromosome ends (Fig. 3b). Given the presence of 28 chromosome ends per haploid blood-stage parasite, this suggests the presence of 7 to 4 chromosome ends per average cluster. Hybridization of the same parasites with the single copy gene *Pf11-1* (located on chromosome 10, about 100 kb proximal to a telomere) showed only one hybridization spot (Fig. 3a). The hybridization spots detected by *rep20* were predominantly located at the nuclear periphery, as shown by the additional 4,6-diamidino-2-phenylindole (DAPI) staining.

Gametocyte sexual blood-stage forms of 3D7 parasites were also subjected to FISH with the *rep20* and *Pf11-1* probes. *Pf11-1* hybridized to a single spot per haploid parasite nucleus (Fig. 3c). With *rep20*, telomere clustering of the chromosome ends was observed in gametocytes (stages III and IV) and resembled the signal seen in asexual blood stages. However, the telomere clusters were found to be concentrated at one pole of the nucleus, forming the image of a classical meiotic bouquet-like structure²¹ in the elongated haploid nuclei (Fig. 3d). We assume that the bouquet-related alignment of telomeres positions non-homologous chromo-

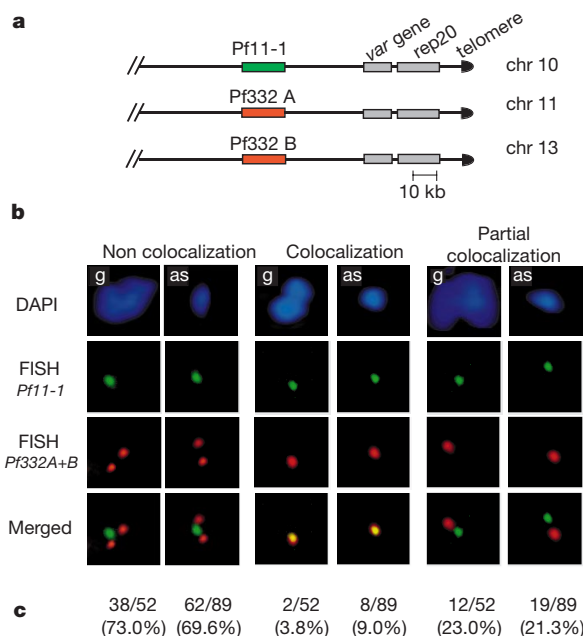


Figure 4 Promiscuous bouquet formation of *P. falciparum* chromosome ends.

a, Representation of subtelomeric regions of HB3 chromosomes 10, 11 and 13, and the gene markers used for FISH analysis. **b**, Colocalization FISH analysis of heterologous telomeres revealed with *Pf11-1*-specific (green) and *Pf332*-specific (red) gene probes. A typical example of an asexual (as) and gametocyte (g) of HB3 parasites is shown for each category. **c**, The frequency of complete, partial or non-colocalization of the three chromosome end specific markers is indicated. The number of nuclei analysed was 89 asexual-stage nuclei and 52 gametocytes.

some strands close together, thus forming a structural framework for ectopic recombination events. The conserved higher-order structure of chromosome ends (Fig. 3e) may promote ectopic exchange of sequences in the subtelomeric regions of *P. falciparum*.

We investigated whether clustering of *P. falciparum* chromosome ends is random or whether the same ends are always involved in cluster formation. We labelled the ends of three distinct chromosomes to determine the frequency of paired heterologous telomeres, using FISH analysis. Two chromosome-specific subtelomeric gene markers were used: *Pf11-1* (adjacent to *HB3var10-1*, data not shown) for chromosome 10; and *Pf332* for chromosome 11 (*Pf332-A*) and chromosome 13 (*Pf332-B*; duplicated in HB3 parasites²³ as shown in Fig. 4a). *Pf332-A* and *Pf332-B* have been shown to undergo unusually high levels of ectopic recombination during meiosis²³. Both subtelomeric markers *Pf11-1* and *Pf332* colocalize with one of the clusters (data not shown). We observe promiscuous cluster formation for the three chromosome ends studied (Fig. 4b, c).

This result implies that telomere clustering in *P. falciparum* asexual and sexual stages appears to have a random component, suggesting that a given subtelomeric *var* gene might potentially recombine with *var* genes from any chromosome end. *P. falciparum* telomere-associated sequences are very well conserved²⁴, and we speculate that this shared chromosome structure is important for promiscuous telomere cluster formation. However, recombination between certain combination of chromosomes ends may be more efficient than between others, depending on *var* sequence homology and chromatin structure.

We have shown that a human pathogen integrates nuclear architecture into its survival strategy to enhance the emergence of parasites with new antigenic and adhesive combinations of *var* genes. *P. falciparum* populations harbour very many *var* forms, whose diversity and continual renewal support the success of this species against host immunity. Telomere-mediated enhancement of recombination helps to generate this diversity and perhaps augments the variation of the relatively small *var* gene complement (approximately 50 genes) in individual parasites during chronic asexual blood-stage infection.

In the absence of selective pressure by the host's immune system in cultured *P. falciparum* parasites, we assume that the occasional occurrence of ectopic recombination in *var* genes will remain undetectable in bulk-cultured parasites. A chromosome-internal *var* gene DNA rearrangement has been reported to occur in laboratory-cultured parasites, resulting in the complete deletion of a *var* gene²⁵. This recombination, however, differs from telomeric recombination as it involved *var* copies arranged *cis* to one another. By contrast, telomere-associated gene conversion involves sequences on different chromosome ends and leads to new *var* forms.

The spatial organization of chromosome ends in the nucleus also facilitates gene silencing in other organisms²⁶, so the association of subtelomeric *var* genes with such 'silent clusters' might affect possible epigenetic mechanisms of *var* gene regulation. In addition, our study sheds light on the evolution of virulence determinants, and other human protozoan pathogens may follow the same blueprint. □

Methods

P. falciparum culture

P. falciparum blood-stage parasites from the HB3×3D7 and HB3×HB3 crosses²⁷ and the HB3×Dd2 cross²⁸ were cultured as described¹². Progeny clones XP1, XP2, XP3, X10, X12, X13, X51 and X45 were kindly provided by D. Walliker. Induction of sexual blood stages of *in vitro* cultured 3D7 and HB3 parasites was done using standard protocols²⁹.

PFGE and chromosome mapping

Pulsed field gel electrophoresis (PFGE) and two-dimensional PFGE (2D-PFGE) was performed as described¹⁵.

Isolation of *var* sequence tags and Southern hybridizations

The *var* sequence tags were obtained using primers varA5'2 and varE3'2 that amplify fragments of approximately 520 bp from the DBL-α domain of most *var* genes in *P. falciparum*^{12,15}. DBL-α sequence was extended using vectorette (Genosys Biotechnologies Inc.). Markers MSA-1 (chromosome 9), *Pf11-1* (chromosome 10) and *Pf332* (chromosomes 11 and 13 in HB3 parasites) are reported elsewhere^{23,30,31}. Southern analysis was carried out as described^{12,15}.

Fluorescent *in situ* hybridization (FISH)

HB3, 3D7, Dd2 and Palo-Alto strain parasites were washed three times in RPMI 1640 (Gibco BRL). Thin blood films were prepared on standard microscope slides. The slides were air dried for 30 min at room temperature followed by fixation in 4% paraformaldehyde/PBS (phosphate buffered saline) for 10 min. The FISH technique was performed as described³². After washing, the slides were mounted in antifade medium and visualized using a Nikon E800 microscope. For colocalization studies, optical section stacks (0.2 µm per section) were collected using a Leica DMRX14 microscope equipped with a Hamamatsu cooled CCD (charge coupled device). Composite images were produced using Photoshop (Adobe Systems).

Received 8 August; accepted 13 September 2000.

- Miller, L. H., Good, M. F. & Milon, G. Malaria pathogenesis. *Science* **264**, 1878–1883 (1994).
- Borst, P., Bitter, W., McCulloch, R., Van Leeuwen, F. & Rudenko, G. Antigenic variation in malaria. *Cell* **82**, 1–4 (1995).
- Biggs, B. A. et al. Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. *J. Immunol.* **149**, 2047–2054 (1992).
- Roberts, D. J. et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357**, 689–692 (1992).
- Carlson, J. et al. Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. *Lancet* **336**, 1457–1460 (1990).
- Fried, M. & Duffy, P. Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. *Science* **272**, 1502–1504 (1996).
- Bull, P. C. et al. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Med.* **4**, 358–360 (1998).
- Baruch, D. I. et al. Cloning of *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* **82**, 77–87 (1995).
- Su, X. Z. et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100 (1995).
- Smith, J. D. et al. Switches in expression of *Plasmodium falciparum* *var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110 (1995).
- Chen, Q. et al. Developmental selection of *var* gene expression in *Plasmodium falciparum*. *Nature* **394**, 392–395 (1998).
- Scherf, A. et al. Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J.* **17**, 5418–5426 (1998).
- Su, X. Z. et al. A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* **286**, 1351–1353 (1999).
- Gupta, S., Snow, R. W., Donnelly, C. & Newbold, C. Acquired immunity and postnatal clinical protection in childhood cerebral malaria. *Proc. R. Soc. Lond. B* **266**, 33–38 (1999).
- Hernandez-Rivas, R. et al. Expressed *var* genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol. Cell. Biol.* **17**, 604–611 (1997).
- Gardner, M. J. et al. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132 (1998).
- Bowman, S. et al. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
- Gupta, S. & Anderson, R. M. Population structure of pathogens: The role of immune selection. *Parasitol. Today* **15**, 497–501 (1999).
- Pryde, F. E., Gorham, H. C. & Louis, E. J. Chromosome ends: all the same under their caps. *Curr. Opin. Genet. Dev.* **7**, 822–8 (1997).
- Dernburg, A. E., Sedat, J. W., Cande, W. Z. & Bass, H. W. In *Telomeres* (eds Blackburn, E. H. & Greider, C. W.) 295–338 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1995).
- de Lange, T. Ending up with the right partner. *Nature* **392**, 753–754 (1998).
- Cooper, J. P., Watanabe, Y. & Nurse, P. Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination. *Nature* **392**, 828–831 (1998).
- Hinterberg, K., Mattei, D., Welles, T. E. & Scherf, A. Interchromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross. *EMBO J.* **13**, 4174–4180 (1994).
- Figueiredo, L. M., Pirrit, L. A. & Scherf, A. Genomic organisation and chromatin structure of *Plasmodium falciparum* chromosome ends. *Mol. Biochem. Parasitol.* **106**, 169–174 (2000).
- Deitsch, K. W., del Pinal, A. & Welles, T. E. Intra-cluster recombination and *var* transcription switches in the antigenic variation of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **101**, 107–116 (1999).
- Andrulis, E. D., Neiman, A. M., Zappulla, D. C. & Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **394**, 592–595 (1998); erratum *Nature* **395**, 525 (1998).
- Walliker, D. et al. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666 (1987).
- Welles, T. E. et al. Chloroquine resistance not linked to *mdr*-like genes in a *Plasmodium falciparum* cross. *Nature* **345**, 253–255 (1990).
- Carter, R., Ranford-Cartwright, L. & Alano, P. The culture and preparation of gametocytes of *Plasmodium falciparum* for immunochemical, molecular, and mosquito infectivity studies. *Methods Mol. Biol.* **21**, 67–88 (1993).
- Scherf, A. et al. The 11-1 gene of *Plasmodium falciparum* codes for distinct fast evolving repeats. *EMBO J.* **7**, 1129–1137 (1988).
- Kimura, E., Mattei, D., di Santi, S. M. & Scherf, A. Genetic diversity in the major merozoite surface

- antigen of *Plasmodium falciparum*: high prevalence of a third polymorphic form detected in strains derived from malaria patients. *Gene* **91**, 57–62 (1990).
32. Ersfeld, K. & Gull, K. Partitioning of large and minichromosomes in *Trypanosoma brucei*. *Science* **276**, 611–614 (1997).

Acknowledgements

We thank C. Roth for critically reading the manuscript, D. Mattei for providing the Pf332 probes, V. Galy for help with the colocalization studies and D. Walliker for the progeny clones of the HB3×3D7 cross. This work has been supported by grants from the Commission of the European Communities for research and technical development. L.H.F.-J. was supported by a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) fellowship. L.A.P. was supported by a European Community grant.

Correspondence and requests for materials should be addressed to A.S.
(e-mail: ascherf@pasteur.fr).

Matrix proteins can generate the higher order architecture of the Golgi apparatus

Joachim Seemann*, Eija Jokitalo†, Marc Pypaert* & Graham Warren*

* Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, PO Box 208002, New Haven, Connecticut 06520-8002, USA

† Institute of Biotechnology, Electron Microscopy Unit, Biocenter 1, PO Box 56 (Viikinkaari 9), 00014 University of Helsinki, Finland

The Golgi apparatus in animal cells comprises a reticulum of linked stacks in the pericentriolar and often in the juxtanuclear regions of the cell¹. The unique architecture of this organelle is thought to depend on the cytoskeleton² and cytoplasmic matrix proteins^{3,4}—the best characterized being the golgin family of fibrous, coiled-coil proteins and the GRASP family of stacking proteins^{5–10}. Here we show that these matrix proteins can be separated from oligosaccharide-modifying enzymes in the Golgi stack without affecting their ability to form a ribbon-like reticulum in the correct location near to the nucleus. Our data suggest that the Golgi is a structural scaffold that can exist independently of, but is normally populated by, the enzyme-containing membranes that modify transiting cargo. This new concept of the Golgi further indicates that the Golgi may be an autonomous organelle rather than one that is in simple dynamic equilibrium with the endoplasmic reticulum.

The primary function of the Golgi apparatus is the stepwise modification and sorting of cargo synthesized in the endoplasmic reticulum (ER) and destined for different cellular and extracellular locations¹¹. The role of the complex architecture of the Golgi in this process has still not been elucidated, but one approach would be to separate those proteins that help determine the structure from those that carry out its function. We have discovered a way to separate Golgi matrix proteins from oligosaccharide-modifying enzymes, by exploiting the properties of Brefeldin A (BFA) and a Sar1 dominant-negative mutant that blocks exit from the ER by preventing the uncoating of COPII (coat protein complex II)-coated vesicles^{12,13}.

Brefeldin A has a marked effect on the Golgi apparatus, the best characterized being the transfer of Golgi enzymes to the lumen of the ER^{14,15}. Less well characterized are the membranes that are left behind, dispersed tubulo-vesicular clusters¹⁶—termed Golgi remnants—that contain recycling proteins such as ERGIC-53 (refs 17, 18).

We previously looked at one of the golgins, GM130, in normal rat kidney (NRK) cells treated with BFA. Instead of finding it in the ER, along with Golgi enzymes such as α -mannosidase II (Man II)

(Fig. 1a), we found it in well-distributed punctate structures (Fig. 1a), which colocalized in part with ERGIC-53 (ref. 5). This suggested that GM130 was present in the Golgi remnants after BFA treatment.

We then injected a dominant-negative form of Sar1 (Sar1^{DN}) to block exit from the ER^{12,13} once the BFA was removed¹⁹. Pilot experiments using the plasma membrane marker, CD8, showed that this mutant prevented the appearance of CD8 on the cell surface.

Removal of BFA in the presence of Sar1^{DN} had the expected effect on the Golgi enzyme, Man II, which was retained in the ER (Fig. 1a). Earlier work had shown that the fate of ERGIC-53 under these conditions was to accumulate in the ER¹⁹. Presumably ERGIC-53 recycles from the Golgi remnants to the ER but cannot subsequently move back out.

In contrast, staining for GM130 changed from a dispersed punctate pattern to one that resembled the Golgi apparatus seen in interphase cells. Figure 1a shows the ribbon-like structure near to the nucleus that contains GM130 but lacks Man II. Similar results were obtained for other matrix proteins, GRASP65, Giantin and p115 (Fig. 1b). All gave the same dispersed pattern of puncta in the presence of BFA, and all reformed a Golgi ribbon after washout in the presence of Sar1^{DN}. Other matrix proteins (for example, GRASP55; data not shown) yielded the same result, whereas other Golgi enzymes, such as β -1,4 galactosyltransferase (GalT), were retained in the ER (Fig. 1b).

We analysed these structures by electron microscopy. Tubulo-vesicular Golgi remnants were observed in Epon sections of NRK cells treated with BFA (Fig. 2a), and these re-formed a typical Golgi ribbon when the BFA was washed out (Fig. 2b). Injection of Sar1^{DN} before washout yielded a striking collection of vesicles and vacuoles, near to the nucleus and the centrioles (Fig. 2c). Some of the vacuoles were interdigitated, suggesting that the GRASP stacking proteins were at least partly functional. To confirm that these vesicles and vacuoles were associated with the ribbon seen by immunofluorescence, we carried out immuno-labelling studies. GM130 is restricted to the *cis*-Golgi (Fig. 3a) and was also found on Golgi remnants (Fig. 3b). To label the micro-injected cells, they were fixed and embedded in LR white resin and then labelled with antibodies to GM130 followed by secondary antibodies coupled to gold. Although the morphological preservation is lower by using this method than by using cryo-electron microscopy, clusters of vesicles and vacuoles could clearly be identified, often near to the nucleus and centrioles, and these were labelled with gold particles (Fig. 3c).

To show that the separation of matrix proteins and Golgi enzymes was not the consequence of using a drug, we took advantage of the slow cycle of enzymes between the Golgi and the ER^{20,21}. Other studies have shown that micro-injection of Sar1^{DN} leads to the slow accumulation of Golgi enzymes in the ER, but these studies did not examine the structural consequences to the Golgi apparatus or the fate of any matrix proteins^{21,22}. As shown in Fig. 4, a 5-h incubation after Sar1^{DN} injection led to the relocation of both Man II and GalT to the ER, whereas the pattern for GRASP65 and GM130 was essentially unchanged. At the electron microscope level, clusters of small vesicles were observed in the Golgi region (Fig. 2d), consistent with the loss of enzyme-containing membranes. The vesicles were much smaller than those observed using a combination of BFA and Sar1^{DN} suggesting that BFA delivers additional Golgi proteins to the ER that do not follow the normal slow recycling pathway.

The experiments described here show that Golgi enzymes can be separated from matrix proteins, and that these retain the information needed to form a ribbon-like structure in the correct location in the cell. Our data indicate that the Golgi remnants identified earlier by electron microscopy may be the substrates for this process, as we have shown them to contain a variety of matrix proteins by fluorescence microscopy and at least one of the golgins, GM130,

by immuno-electron microscopy. The function of the *trans*-Golgi network in this process is still unclear. Treatment with BFA causes the *trans*-Golgi network to collapse around the centrosomes^{23,24}, and removal leads to its association with the ribbon-containing matrix proteins. They do not, however, colocalize as would be expected, because the matrix proteins studied here, which are the best characterized, are located in the Golgi stack. This indicates that

the stack matrix proteins may be capable of forming a ribbon on their own.

The membranes that associate with this matrix ribbon have still to be characterized. They lack Golgi enzymes and their morphology depends on whether BFA is part of the process that generates them. Most notably, in the absence of BFA, the membranes appear as small vesicles that are mostly discrete and separated from each other. This

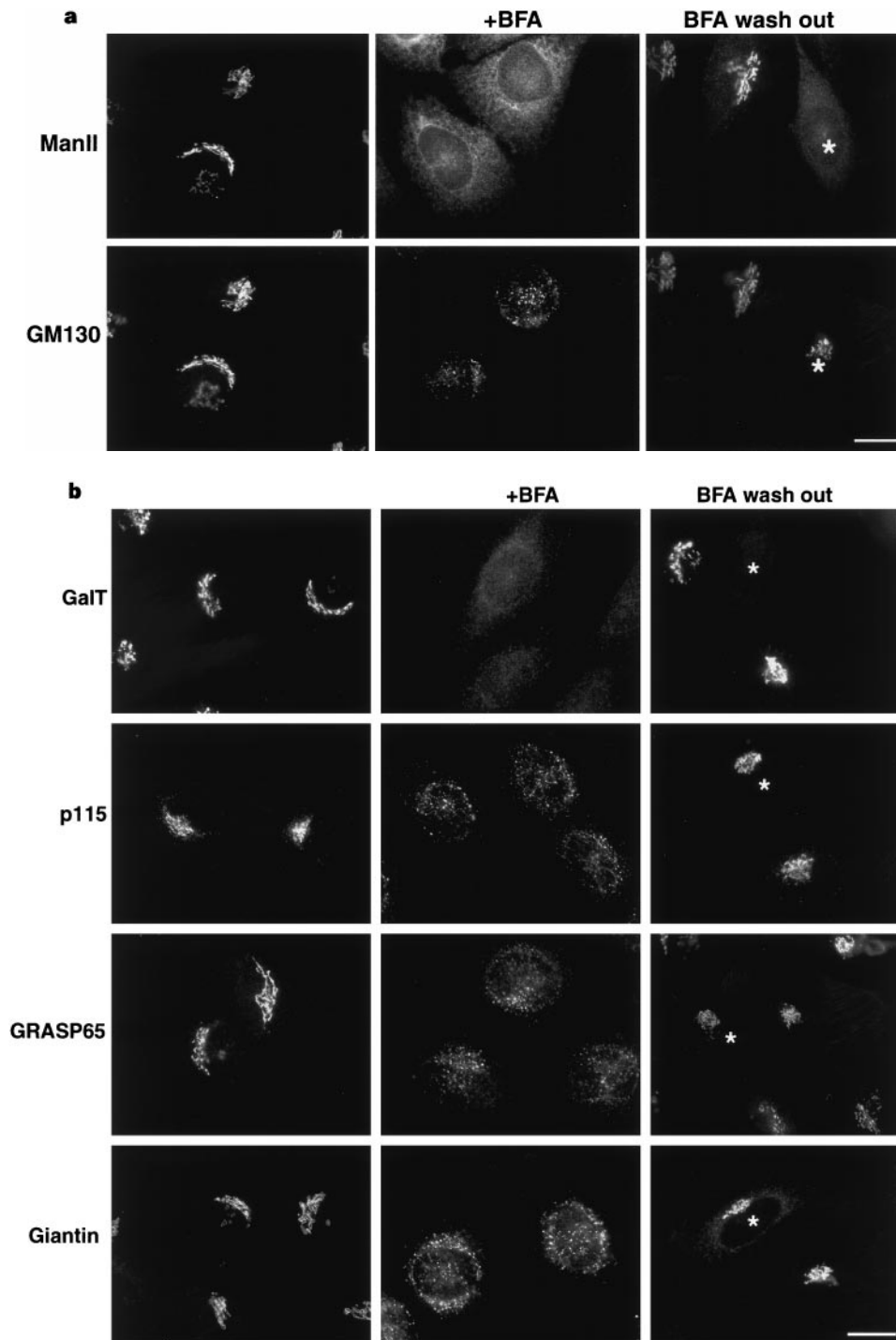


Figure 1 The effect of blocking ER exit by Sar1^{DN} on reassembly of the Golgi apparatus after BFA wash out. NRK cells (left panels) were treated with BFA for 30 min (+BFA, middle panels) or micro-injected with Sar1^{DN} (asterisks) and incubated in the presence of BFA for 30 min. BFA was washed out and the cells were fixed 2 h later (BFA wash out, right panels). **a**, Cells were double labelled with antibodies to Man II and GM130, followed by

appropriate secondary antibodies conjugated to Alexa fluorophors. **b**, Cells were singly labelled with antibodies to the proteins indicated on the left followed by fluorescent secondary antibodies. Micro-injected cells are marked by asterisks. Note that cells injected with Sar1^{DN} have a Golgi ribbon lacking Man II and GalT but containing all the matrix proteins. Bar, 15 μ m.

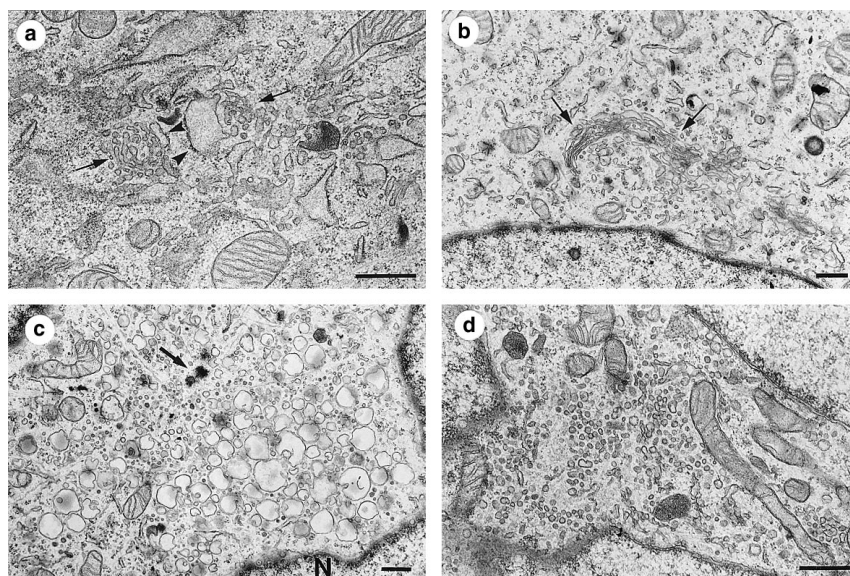


Figure 2 Ultrastructure of NRK cells treated with BFA and/or micro-injected with Sar1^{DN}. **a**, Cells incubated for 30 mins with BFA contained vesiculo-tubular clusters, termed Golgi remnants (arrows), in close proximity to the ER, which was dilated (arrowheads). **b**, After 2 h incubation in the absence of BFA, normal Golgi stacks reformed (arrows). **c**, Injection of Sar1^{DN} into BFA-treated cells before incubation for 2 h in the absence of BFA led to

collections of small vesicles and large vacuoles near the nucleus (N) and centrioles (arrow). **d**, Injection of untreated NRK cells with Sar1^{DN} and incubation for 5 h led to depletion of membranes in the Golgi region, leaving collections of small, dispersed vesicles. Bar, 0.5 μm.

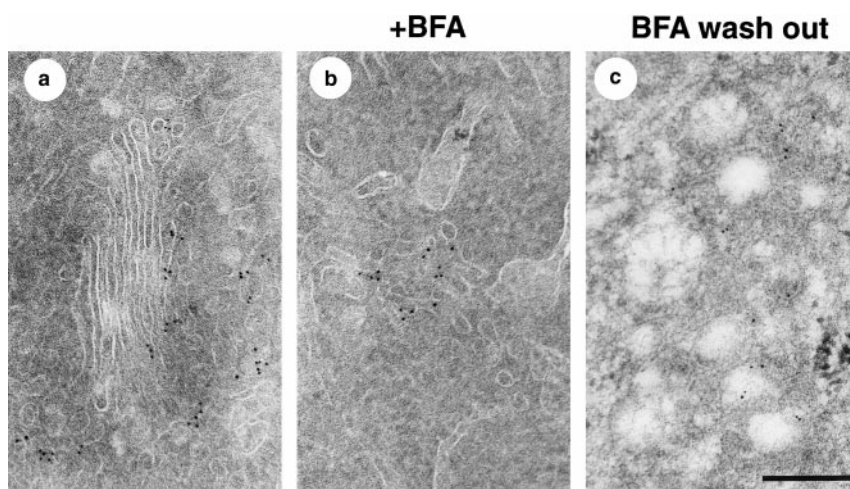


Figure 3 Localization of GM130 by immuno-electron microscopy. NRK cells were processed for cryo-electron microscopy (**a**, **b**) or embedded in LR white (**c**), and then labelled with monoclonal antibodies to GM130 followed by secondary antibodies and protein-A/gold. **a**, In untreated cells GM130 is localized to *cis*-Golgi cisternae. **b**, In BFA-

treated cells, GM130 is found on the tubulo-vesicular Golgi remnants. **c**, In BFA-treated cells injected with Sar1^{DN} and then incubated in the absence of BFA, GM130 was found mostly in the collection of vacuoles near the nucleus and centrioles. Bar, 0.2 μm.

indicates that the membranes may not have a direct structural function in the formation of the ribbon, although they may be responsible for the structure of individual cisternae^{25,26}. This in turn suggests a view of Golgi architecture that focuses on matrix proteins rather than membranes. For example, the properties of the GRASP and golgin family of proteins fit them for the task of forming extensive meshworks that might organize Golgi membranes, binding to them directly or indirectly through lipid anchors. The Golgi apparatus might therefore be viewed as an independent meshwork of proteins that could act as a scaffold for arranging the enzyme-containing membranes that are responsible for carrying out the known functions of the Golgi apparatus.

This view contrasts with that put forward by other workers, who

view the Golgi as an organelle in dynamic equilibrium with the ER. During interphase, perturbations in this equilibrium have been proposed to be the means of moving the Golgi from the cell centre to the periphery, through the ER^{20,21}. During mitosis, these same perturbations have been suggested to lead to fusion of the Golgi with the ER so that partitioning of the ER leads to partitioning of the Golgi²⁷. These experiments were, however, conducted using Golgi enzymes as the markers. If the Golgi comprises a separable population of enzymes and matrix proteins, the latter having an existence independent of the ER, we may need to view the Golgi as an autonomous organelle, and one, perhaps, that is based on a structural framework that is normally populated by enzyme-containing membranes. □

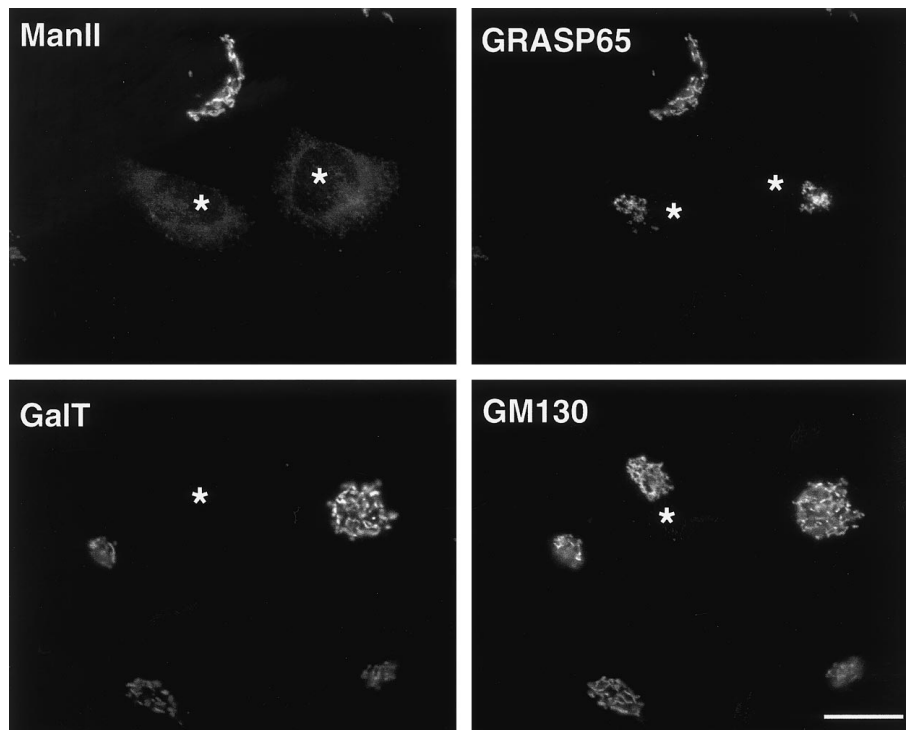


Figure 4 Effect of Sar1^{DN} on the pattern of Golgi enzymes and matrix proteins. NRK cells were micro-injected with Sar1^{DN} and incubated for 5 h before double labelling for Man II/GRASP65 or GalT/GM130. Micro-injected cells are marked by asterisks. Note that both

Man II and GalT are depleted from the Golgi apparatus, whereas the staining patterns for GRASP65 and GM130 appear to be very similar to uninjected cells. Bar, 15 μ m.

Methods

Cell lines and antibodies

We grew normal rat kidney in DMEM medium (Gibco BRL) with 10% FCS at 37°C and 10%. We used the following antibodies (provided by): mAb NN2c10 against GM130 (N. Nakamura); mAb 5F10 and rabbit antibodies MLO-7 against GM130 (M. Lowe); mAb 4H1 against p115 (M. G. Waters); mAb GTL2 against GalT (T. Saganuma); mAb 7E10 against GRASP65 (F. Barr); mAb against CD8 (F. Watt); mAb against ERGIC53 (H. P. Hauri); rabbit antibodies against Giantin (M. Renz); and rabbit antibodies against Man II (T. Moremen). We obtained Alexa Fluor 488 and Alexa Fluor 594 conjugated secondary antibodies from Molecular Probes.

Purification of Sar1

Plasmids encoding mammalian Sar1 mutants were provided by W. E. Balch. Proteins were purified essentially as described^{12,28}. Briefly, His-tagged proteins were expressed and isolated from bacteria using Ni-NTA Agarose (Qiagen), dialysed against 25 mM HEPES pH 7.2, 125 mM potassium acetate, 1 mM MgCl₂, 1 mM glutathione, 10 μ M GDP and 50 μ M EGTA, and snap frozen.

Micro-injection

Micro-injection was performed as described²⁹. We carried out protein injections in the presence of 0.1 mg ml⁻¹ cycloheximide (Sigma), and in all following incubations we added cycloheximide to the medium. Sar1 was injected into the cytoplasm at an initial protein concentration of 0.7–1 mg ml⁻¹, together with the following injection markers: cascade-blue conjugated BSA (Molecular Probes) for immunofluorescence; 10 nm protein A-gold (Department of Cell Biology, Utrecht School of Medicine) for electron microscopy. The cells were either fixed after 5 h of incubation or treated for 30 min with 5 μ g ml⁻¹ Brefeldin A (Epicenter) to disassemble the Golgi apparatus, washed several times and incubated in fresh medium for 120 min. We analysed transport of CD8 as described¹⁹.

Immunofluorescence

Immunofluorescence was carried out as described²⁹. Images were captured using an Orca 100 CCD camera (Hamamatsu) and analysed using Openlab 2.1 (Improvision).

Electron microscopy

We processed micro-injected cells for Epon sections as described²⁹. For immunolabelling, injected cells were fixed with 4% PFA (Electron Microscopy Sciences (EMS)), 0.1% GA (EMS) and 3% sucrose in 0.1 M Na cacodylate pH 7.4, and washed, dehydrated in ethanol and embedded in London resin (LR) white resin (EMS). We cut 60-nm sections parallel to the coverslip using an Ultracut E ultramicrotome (Leica). We performed immunolabelling

of GM130 using mAb NN2C10, followed by rabbit anti-mouse IgG (Cappel) and 5 nm of protein-A/gold (Department of Cell Biology, Utrecht School of Medicine). We air-dried and stained sections with 5% uranyl acetate (EMS) and lead citrate.

For cryosectioning, cells were fixed with 4% PFA, 0.1% GA, 0.25 M HEPES pH 7.4, embedded in 10% gelatin, infiltrated with 2.3 M sucrose in PBS and frozen in liquid nitrogen. We cut 95-nm sections using an Ultracut UCT ultramicrotome with a cryo-attachment (Leica). Immunolabelling was performed as for LR white sections. We stained sections with 2% neutral uranyl acetate, then incubated them with a mixture of 1.8% methyl cellulose and 0.5% uranyl acetate and air-dried. We examined samples using a Philips 410 electron microscope.

Received 10 July; accepted 21 August 2000.

- Farquhar, M. G. & Palade, G. E. The Golgi-apparatus—100 years of progress and controversy. *Trends Cell Biol.* **8**, 2–10 (1998).
- Donaldson, J. G. & Lippincott-Schwartz, J. Sorting and signaling at the Golgi complex. *Cell* **101**, 693–696 (2000).
- Dematteis, M. A. & Morrow, J. S. The role of ankyrin and spectrin in membrane-transport and domain formation. *Curr. Opin. Cell Biol.* **10**, 542–549 (1998).
- Linstedt, A. D. Stacking the cisternae. *Curr. Biol.* **9**, R893–R896 (1999).
- Nakamura, N. *et al.* Characterization of a cis-Golgi matrix protein, GM130. *J. Cell Biol.* **131**, 1715–1726 (1995).
- Barr, F. A. A novel Rab6-interacting domain defines a family of Golgi-targeted coiled-coil proteins. *Curr. Biol.* **9**, 381–384 (1999).
- Munro, S. & Nichols, B. J. The GRIP domain—a novel Golgi-targeting domain found in several coiled-coil proteins. *Curr. Biol.* **9**, 377–380 (1999).
- Shorter, J. *et al.* GRASP55, a second mammalian GRASP protein involved in the stacking of Golgi cisternae in a cell-free system. *EMBO J.* **18**, 4949–4960 (1999).
- Sonnichsen, B. *et al.* Role for Giantin in docking COPI vesicles to Golgi membranes. *J. Cell Biol.* **140**, 1013–1021 (1998).
- Kjer-Nielsen, L., Teasdale, R. D., van Vliet, C. & Gleeson, P. A. A novel Golgi-localisation domain shared by a class of coiled-coil peripheral membrane proteins. *Curr. Biol.* **9**, 385–388 (1999).
- Rothman, J. E. & Wieland, F. T. Protein sorting by transport vesicles. *Science* **272**, 227–234 (1996).
- Aridor, M., Bannykh, S. I., Rowe, T. & Balch, W. E. Sequential coupling between COPII and COPI vesicle coats in endoplasmic-reticulum to Golgi transport. *J. Cell Biol.* **131**, 875–893 (1995).
- Pepperkok, R., Lowe, M., Burke, B. & Kreis, T. E. Three distinct steps in transport of vesicular stomatitis virus glycoprotein from the ER to the cell surface *in vivo* with differential sensitivities to GTP gamma S. *J. Cell Sci.* **111**, 1877–1888 (1998).
- Lippincott-Schwartz, J., Yuan, L. C., Bonifacino, J. S. & Klausner, R. D. Rapid redistribution of Golgi proteins into the ER in cells treated with brefeldin A: evidence for membrane cycling from Golgi to ER. *Cell* **56**, 801–813 (1989).
- Doms, R. W., Russ, G. & Yewdell, J. W. Brefeldin A redistributes resident and itinerant Golgi proteins to the endoplasmic reticulum. *J. Cell Biol.* **109**, 61–72 (1989).
- Sandvig, K., Prydz, K., Hansen, S. H. & van, D. B. Ricin transport in brefeldin A-treated cells: correlation between Golgi structure and toxic effect. *J. Cell Biol.* **115**, 971–981 (1991).

17. Hendricks, L. C., McClanahan, S. L., McCaffrey, M., Palade, G. E. & Farquhar, M. G. Golgi proteins persist in the tubulovesicular remnants found in Brefeldin A-treated pancreatic acinar cells. *Eur. J. Cell Biol.* **58**, 202–213 (1992).
18. Lippincott-Schwartz, J. *et al.* Microtubule-dependent retrograde transport of proteins into the ER in the presence of brefeldin A suggests an ER recycling pathway. *Cell* **60**, 821–836 (1990).
19. Shima, D. T., Cabrera-Poch, N., Pepperkok, R. & Warren, G. An ordered inheritance strategy for the Golgi apparatus: visualization of mitotic disassembly reveals a role for the mitotic spindle. *J. Cell Biol.* **141**, 955–966 (1998).
20. Cole, N. B., Sciaky, N., Marotta, A., Song, J. & Lippincott-Schwartz, J. Golgi dispersal during microtubule disruption—regeneration of Golgi stacks at peripheral endoplasmic-reticulum exit sites. *Mol. Biol. Cell* **7**, 631–650 (1996).
21. Storrie, B. *et al.* Recycling of Golgi-resident glycosyltransferases through the ER reveals a novel pathway and provides an explanation for nocodazole-induced Golgi scattering. *J. Cell Biol.* **143**, 1505–1521 (1998).
22. Girod, A. *et al.* Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum. *Nature Cell Biol.* **1**, 423–430 (1999).
23. Lippincott-Schwartz, J. *et al.* Brefeldin A's effects on endosomes, lysosomes, and the TGN suggest a general mechanism for regulating organelle structure and membrane traffic. *Cell* **67**, 601–616 (1991).
24. Reaves, B. & Banting, G. Perturbation of the morphology of the trans-Golgi network following Brefeldin A treatment: redistribution of a TGN-specific integral membrane protein, TGN38. *J. Cell Biol.* **116**, 85–94 (1992).
25. Cluett, E. B. & Brown, W. J. Adhesion of Golgi cisternae by proteinaceous interactions: intercisternal bridges as putative adhesive structures. *J. Cell Sci.* **103**, 773–784 (1992).
26. Nilsson, T. *et al.* Kin recognition between medial Golgi enzymes in HeLa cells. *EMBO J.* **13**, 562–574 (1994).
27. Zaal, K. J. *et al.* Golgi membranes are absorbed into and reemerge from the ER during mitosis. *Cell* **99**, 589–601 (1999).
28. Rowe, T. & Balch, W. E. Expression and purification of mammalian Sar1. *Methods Enzymol.* **257**, 49–53 (1995).
29. Seemann, J., Jokitalo, E. J. & Warren, G. The role of the tethering proteins p115 and GM130 in transport through the Golgi apparatus *in vivo*. *Mol. Biol. Cell* **11**, 635–645 (2000).

Acknowledgements

We would like to thank J. Shorter for his comments. This work was funded by the NIH and J.S. was supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to G.W. (e-mail: graham.warren@yale.edu).

Steps and fluctuations of *Listeria monocytogenes* during actin-based motility

Scot C. Kuo & James L. McGrath

Department of Biomedical Engineering, The Johns Hopkins University, 720 Rutland Avenue, Baltimore, Maryland 21205, USA

The actin-based motility of the bacterium, *Listeria monocytogenes*, is a model system for understanding motile cell functions involving actin polymerization¹. Although the biochemical and genetic aspects of *Listeria* motility have been intensely studied^{2–5}, biophysical data are sparse⁶. Here we have used high-resolution laser tracking to follow the trailing ends of *Listeria* moving in the lamellae of COS7 cells. We found that pauses during motility occur frequently and that episodes of step-like motion often show pauses spaced at about 5.4 nm, which corresponds to the spatial periodicity of F-actin⁷. We occasionally observed smaller steps (<3 nm), as well as periods of motion with no obvious pauses. Clearly, bacteria do not sense cytoplasmic viscoelasticity because they fluctuate 20 times less than adjacent lipid droplets. Instead, bacteria bind their own actin 'tails', and the anchoring proteins can 'step' along growing filaments within the actin tail. Because positional fluctuations are unusually small, the forces of association and propulsion must be very strong. Our data disprove the brownian ratchet model⁸ and limit alternative models, such as the 'elastic' brownian ratchet⁹ or the 'molecular' ratchet^{4,10}.

To monitor the motions of *Listeria* in living cells, we adapted a high-resolution laser-based device originally developed for spherical particles in optical tweezers. Such devices use split photodiode detectors to monitor forward-scattered laser light, often from the trapping laser^{11–15}. Our device avoids optical forces by using extremely low optical power (<0.15 mW) and uses a quadrant photodiode detector to provide two-dimensional (2D) tracking. This device provided the basis of a new approach, called laser-tracking microrheology, to measure the viscoelastic moduli of polymeric solutions and of subcellular regions in living cells^{16–18}. To adapt the device for rod-shaped bacteria, we mapped the optical response while performing a 2D raster scan of fixed bacteria (Fig. 1). Unlike spherical particles, the position of highest signal sensitivity was not at the centre of bacteria, but at the centre of their hemispherical ends. Furthermore, laser-tracking signals were non-linear with position and depended on bacterial orientation relative to the axes of the quadrant photodiode detector. To account for these effects, we perform automated 2D raster calibrations with the piezostage before every high-resolution acquisition.

Data from tracking the actin-based motility of *Listeria* in lamellae of living COS7 cells are shown in Fig. 2. By automated repositioning of the piezostage, we followed the trailing end of a bacterium for 2 min and acquired high-resolution data at each piezostage position (Fig. 2b). Two features are apparent from these trajectories (Fig. 2c). First, bacterial fluctuations are very small; across all 20 high-resolution trajectories, bacterial fluctuations are 1.31 ± 0.04 nm and 0.94 ± 0.03 nm (mean r.m.s. $\pm$ s.e.m.) in the directions parallel and perpendicular to movement, respectively. Second, pauses greater than 50 ms are common during motility and appear in 85% of these 1-s trajectories. Despite hundreds of filaments and fast motility, episodes of step-like progress are common and pauses are often spaced about 5.4 nm apart (Fig. 2c). These steps are considerably smaller than the micrometre-scaled episodic motion observed in certain *Listeria* mutants^{3,19}. Spatial frequencies can be analysed using histograms of all pairwise displacements²⁰ followed by Fourier transform analysis of histograms¹² (Fig. 2d, e). Such analysis of the data in Fig. 2c shows only spatial frequencies of 0.18 nm^{-1} (5.5 nm) and no evidence for higher harmonics (for example, 2.7 nm). In other trajectories, a spacing of 5–6 nm between pauses is dominant in positional histograms, although smaller steps (~ 3 nm) are

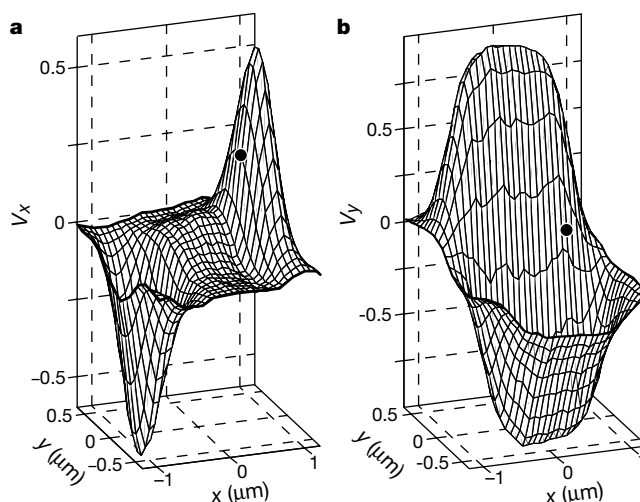


Figure 1 Response of laser-tracking signals with bacterial position. Dead bacteria were adsorbed to glass coverslips and scanned with a 2D raster pattern in 75-nm steps with the piezostage. Voltage signals are shown parallel (V_x) (a) and perpendicular (V_y) (b) to the bacterium's long axis. The dot represents unique (V_x , V_y) signals that would be used by software to position the piezostage to track motion in the $-x$ direction. For live-cell tracking, calibration patterns are much smaller (see Methods).

detectable. These features persist in instances of faster motility (see the $7 \mu\text{m min}^{-1}$ example in Supplementary Information). In general, noise in positional measurements ($1.9 \pm 0.9 \text{ nm}$, mean r.m.s. $\pm$ s.d.) can obscure short-lived steps of about 2.5 nm that may occur in episodes of apparently continuous movement. The $\sim 5.4\text{-nm}$ spacing corresponds to the size of G-actin and to the spatial periodicity of F-actin⁷ (the significance of which will be discussed below).

Because fluctuations can reveal local mechanical properties, we show that fluctuations in positional measurements are largely due to the optical fluctuations of surrounding lamellae (Fig. 3). On exposure to *Listeria*, all cells show increased optical fluctuations in their lamellae, even when bacteria are not internalized. This may be due to bacterial secretion of listeriolysin O, a pore-forming cytolysin, into the culture media (D. A. Portnoy, personal communication). To estimate an upper bound for the magnitude of brownian fluctuations of *Listeria*, we note that all fluctuation measurements in Fig. 3 have standard errors of less than 0.04 nm ($\sim 3\%$). To be indistinguishable ($P > 0.95$) from the interfering background signals, the addition of *Listeria* fluctuations must be within two standard errors. Hence, an initial estimate of *Listeria* fluctuations is less than 0.1 nm ($2.5 \times \text{s.e.m.}$), which is smaller than a covalent bond. Clearly, molecules on the surface of moving *Listeria* can fluctuate much more than the whole bacterium.

No comparable organelle in COS7 shows such small fluctuations. Because they are relatively rigid and distribute throughout COS7 cells, lipid droplets make ideal probes for comparing subcellular mechanics¹⁷. Lipid droplets fluctuate the least (9.9 nm) in lamellae, making lamellae the most rigid region of cytoplasm¹⁷. Even when compensating for their larger size and shape, motile *Listeria* fluctuate at least 20-fold less than is expected from lipid droplets near the bacterium's path (Fig. 4). Because bacteria do not 'sense' the local cytoplasmic viscoelasticity, they must be tightly associated to their actin 'tails'. From studies of reconstituted *Listeria* motility in extracts, theoretical implications of the medium's low viscosity²¹ and direct 'plucking' with optical tweezers²² also argue for tail binding by bacteria. By the equipartition principle²³, the small brownian fluctuations of bacteria ($\langle \Delta x^2 \rangle < 0.1 \text{ nm}^2$) indicate strong tethering forces with an apparent spring constant, $\kappa = k_B T / \langle \Delta x^2 \rangle > 40 \text{ pN nm}^{-1}$, where k_B is Boltzmann's constant and T is absolute temperature. If such tail-binding remained

constant, then more than 220 pN of force ($F = \kappa s$) would be needed to make a step ($s = 5.4 \text{ nm}$).

The brownian ratchet model⁸ makes specific predictions of bacterial fluctuations. In this model, F-actin does not flex and positional fluctuations of bacteria must be large enough to allow intercalation of G-actin ($> 2.7 \text{ nm}$) at appropriate timescales. To evaluate fluctuations, we calculate a mean-squared residual-displacement (MSrD; see Methods) where linear trends are removed before statistical analysis. Like the more traditional mean-squared displacement (MSD) analysis²⁴, MSrD measures the extent of diffusive motion at different timescales. Unlike the MSD, however, MSrD analyses of the brownian ratchet model are not quadratic functions of lag time (τ), but are generally linear functions that show the magnitude of diffusion. At longer lag times ($\tau \approx T/2$, where T is the duration of each simulated trajectory), MSrD(τ) reaches a maximum because numerical detrending necessarily removes the slower diffusive components of the brownian ratchet. From MSrD(τ) analysis of *Listeria* motion (Fig. 4), it is clear that the diffusive component of brownian ratchet motions should have been detectable despite the large optical fluctuations of the lamellae. Compared with velocity-matched simulations of the brownian ratchet ($1.2 \mu\text{m min}^{-1}$ for Fig. 4), fluctuations of bacteria are markedly different ($P < 0.005$).

These data directly contradict the brownian ratchet model in four ways. First, fluctuations of bacteria are much smaller than the intercalation size of G-actin. Our best estimates are less than 0.1 nm . Second, MSrD(τ) analysis suggests that bacteria do not diffuse according to predictions from brownian ratchet simulations. Third, because bacteria fluctuate 20 times less than adjacent particles, bacteria bind their actin 'tails'. Indeed, when motility occasionally stops, bacteria can detach from their tails and show extremely large fluctuations, including 'fish-tailing' (see movie in Supplementary Information). Last, because *Listeria* have episodes of motility with pauses spaced at about 5.4 nm , the bacteria probably step along growing actin filaments. Because hundreds of filaments are in the actin tail, it is remarkable that these episodes of step-like progress occur. Either progress along many filaments is coordinated, or attachment to only a few filaments, perhaps those stretched under the greatest tension, limits bacterial motility.

The discovery that *Listeria* can make steps at the spatial periodicity of F-actin critically modifies remaining models of *Listeria*

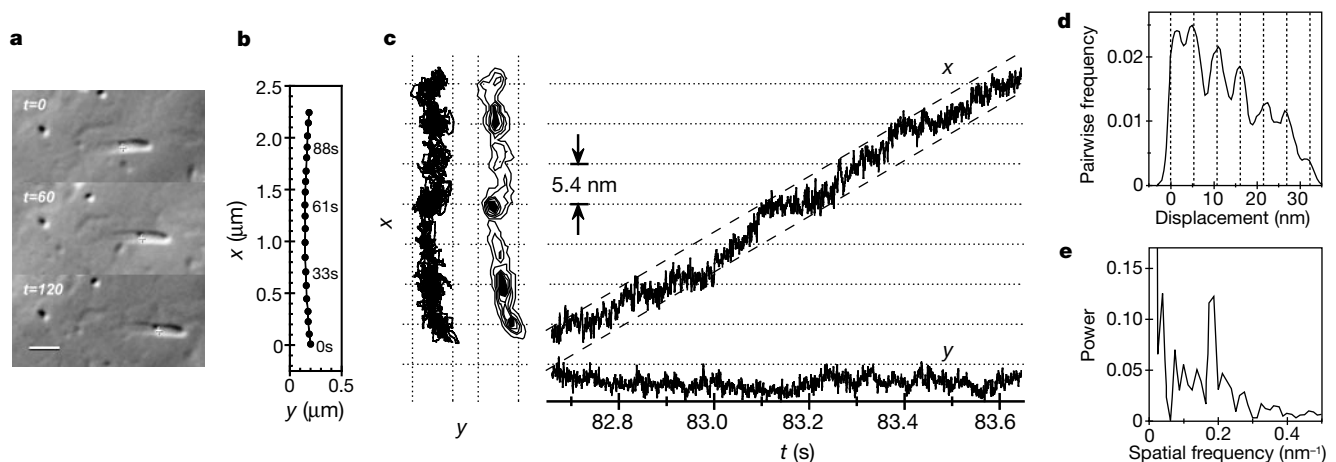


Figure 2 Actin-based motility in COS7 lamellae. **a**, Video-enhanced differential-interference contrast images of motility (laser focused at crosses; scale bar, $2 \mu\text{m}$). **b**, Piezostage coordinates for each high-resolution acquisition (overall speed $1.2 \mu\text{m min}^{-1}$ along x). **c**, High-resolution trajectory with many pauses (x , r.m.s. 1.5 nm ; linear-fit slope $2.3 \mu\text{m min}^{-1}$). Left, the bacterium's trajectory and a contour plot of its positional frequencies (bins $0.75 \times 0.75 \text{ nm}$; local maxima $> 20 \text{ ms nm}^{-2}$ filled black) with transposed axes. Right, bacterial position versus time. All grid lines are spaced at

5.4 nm , and dashed lines denote the intercalation size for G-actin (linear fit $\pm 2.7 \text{ nm}$). **d**, **e**, Spatial frequency analysis of step size. **d**, Normalized histogram of pairwise displacements. For the positional data of **c**, histogram (bins 0.5 nm) of displacements between all pairs of coordinates show a strong 5.4-nm periodicity (grid spacing 5.4 nm). **e**, Fourier analysis (power spectrum) of pairwise displacement histogram in **d** show only 0.18 nm^{-1} (5.5 nm) and no higher spatial frequencies.

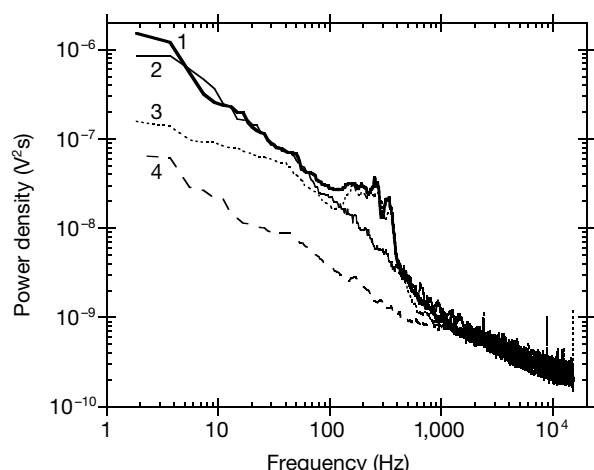


Figure 3 Power spectra of laser-tracking signals. Voltage fluctuations of moving *Listeria* (1); empty lamellae within 2 μm of moving bacterium (2); dead bacterium (0.02% azide) embedded in 15% polyacrylamide (3); and cell-free region of specimen (4). Before calculating power spectra, all voltage data were detrended by subtracting any linear moving component. Mechanical resonances (~ 250 Hz) of the piezostage and microscope are only visible with bacteria. Power spectra from several acquisitions were averaged for this figure, using $n = 40, 14, 49$ and 56 for spectra 1–4, respectively.

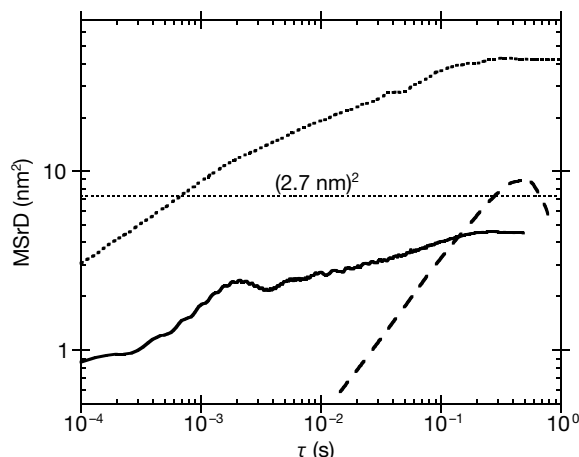


Figure 4 Comparisons of fluctuations. Mean-squared residual displacement (MSrD) was calculated as described in Methods for the motions of adjacent lipid droplet (dotted line); simulation of the brownian ratchet model (dashed line); and actin-based motility of *Listeria* averaged across all acquisitions of Fig. 2b (solid line). Model simulations matched the duration of *Listeria* data sets. For comparison, the intercalation gap for G-actin, 2.7 nm, is drawn as a horizontal, thin dotted line ($\text{MSrD} = 7.3 \text{ nm}^2$).

motility. In the ‘elastic’ brownian ratchet model⁹, bacteria do not bind F-actin and the bending fluctuations of actin filaments allow intercalation of G-actin. From our data, bacteria bind their actin tails and the binding is very restrictive. To observe bacterial steps at the periodicity of F-actin, the ends of bound filaments must fluctuate less than the size of G-actin, hence preventing monomer intercalation by this model. In the ‘molecular’ ratchet model^{4,10}, filament growth occurs when the barbed end of F-actin cyclically attaches and detaches from bacterium-bound protein complexes. For the actin-based motility of *Listeria* and *Shigella flexneri*, F-actin binding is mediated by vasodilator-stimulated phosphoprotein (VASP) bound to ActA or neuronal Wiskott–Aldrich syndrome protein (N-WASP) bound to IcsA, respectively^{2–4,10}. In the molecular ratchet, tethered filaments are under compression when growing. Just as observed in actomyosin experiments^{25–27}, however, F-actin will bend under compression, hence shortening, if not obscuring, steps. Many hundred filaments about *Listeria*, and different elongation mechanisms may operate on different filaments.

To show steps at the spatial periodicity of F-actin, bacterial motion is most probably limited by binding to a few filaments that are stretched taut. Unless filaments undergo coordinated growth, longer filaments will be flexed providing ‘pressure’ and a few tethered filaments become stretched (see Supplementary Information). Probably tethered to the sides of actin filaments by the VASP–ActA protein complex, bacteria ‘slip’ along these taut filaments to reveal their periodicity of 5.4 nm. If several staggered filaments limit bacterial motility, steps could be smaller and have uneven spacing. Reminiscent of molecular motors, stepping requires some processivity of the tethering proteins along the taut filaments as they grow. Otherwise, long periods of dissociation would skip or obscure steps, particularly when taut filaments relax. Analogies to molecular motors will probably guide future biophysical studies of the actin-based motility of *Listeria*. □

Methods

Host cells and infections

We tried using J774 macrophages as hosts for *Listeria* infections²⁸, but J774 cells spread poorly. Thick cells were rich with organelles that confused laser-tracking signals, and the minority of well-spread J774 cells seemed to resist *Listeria* infection. We therefore used the kidney-derived COS7 cell line (ATCC CRL-1651) in which we previously measured subcellular mechanical properties¹⁷. Although COS7 cells are not as readily infected as J774 macrophages, *Listeria* can infect cells with well-spread lamellae and can move on actin-rich tails that stain with fluorescent phalloidin (data not shown).

We grew and maintained virulent *L. monocytogenes* (strain 10403S)²⁸ and COS7 cells¹⁷ as described. We modified these protocols for infection and microscopy by using CO_2 -independent Leibovitz’s L15 medium (Life Technologies) supplemented with 10% fetal bovine serum. After 1 h of infection (multiplicity of infection, MOI = 370) in this medium, excess bacteria were removed by extensive washing and infected cells were allowed to grow for 3 h in medium supplemented with $5 \mu\text{g ml}^{-1}$ gentamicin to kill free-swimming bacteria²⁸. We carried out all infections and growth at 37°C , and all microscopy and laser tracking at room temperature (23°C).

Laser tracking and data analysis

During experiments, automated software cyclically repeated the following sequence: trailing ends of bacteria were placed at the laser focus (Fig. 1) by moving the piezostage (Queensgate Instruments); laser-tracking signals were calibrated with a 2D raster scan spanning 240 nm in 30-nm steps; and high-resolution position data (30 kHz) were recorded with the piezostage stationary. Except for power spectra (Fig. 3), we reduced electro-optical noise by digital filtering with a 5-kHz cut-off, subsampled at 10 kHz, and median filtered in a sliding window of 0.9 ms. Using corresponding raster calibrations, 2D interpolation (MatLab; MathWorks) converted laser-tracking signals into positions. After removing the linear trend in a trajectory $x(t)$, we calculated r.m.s. values by $\text{r.m.s.} = \sqrt{(\sum x^2/N)}$ for N positions over the whole duration, T , of the high-resolution trajectory. Also starting with detrended $x(t)$, we calculated the mean-squared residual displacement (MSrD) as a function of lag time, τ , by $\text{MSrD}(\tau) = \sum [x(t + \tau) - x(t)]^2/n$ for all possible n pairs of positions ($n < N$). For spatial frequency analysis (Fig. 2d), trajectories were first median filtered with a window of 3.5 ms. We calculated displacements as $(x(t_2) - x(t_1))$ for all data pairs $t_2 > t_1 \geq 0$, and histograms of these displacements were normalized by the total number of pairwise displacements, $N(N-1)/2$ (ref. 20). We calculated the one-sided power spectrum (Fig. 2e) from the pairwise displacements histogram, with negative displacements excluded¹². For Fig. 4, simulations of the ideal brownian ratchet used step size $\delta = 2.7$ nm and diffusion coefficient $D = 28 \text{ nm}^2 \text{ s}^{-1}$ so that the average velocity is $1.2 \mu\text{m min}^{-1}$ (ref. 8). We calculated the simulation with 0.2-ms temporal resolution, and we calculated MSrD as described for moving *Listeria*.

Received 20 April; accepted 22 August 2000.

- Borisov, G. G. & Svitkina, T. M. Actin machinery: pushing the envelope. *Curr. Opin. Cell Biol.* **12**, 104–112 (2000).
- Smith, G. A., Theriot, J. A. & Portnoy, D. A. The tandem repeat domain in the *Listeria monocytogenes* ActA protein controls the rate of actin-based motility, the percentage of moving bacteria, and the localization of vasodilator-stimulated phosphoprotein and profilin. *J. Cell Biol.* **135**, 647–660 (1996).
- Lasa, I. et al. Identification of two regions in the N-terminal domain of ActA involved in the actin comet tail formation by *Listeria monocytogenes*. *EMBO J.* **16**, 1531–1540 (1997).
- Laurent, B. et al. Role of proteins of the Ena/VASP family in actin-based motility of *Listeria monocytogenes*. *J. Cell Biol.* **144**, 1245–1258 (1999).
- Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–615 (1999).
- Theriot, J. A. The polymerization motor. *Traffic* **1**, 19–28 (2000).
- Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
- Peskin, C., Odell, G. & Oster, G. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J.* **65**, 316–324 (1993).
- Mogilner, A. & Oster, G. Cell motility driven by actin polymerization. *Biophys. J.* **71**, 3030–3045 (1996).
- Egile, C. et al. Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. *J. Cell Biol.* **146**, 1319–1332 (1999).

11. Denk, W., Webb, W. W. & Hudspeeth, A. J. Mechanical properties of sensory hair bundles are reflected in their Brownian motion measured with a laser differential interferometer. *Proc. Natl Acad. Sci. USA* **86**, 5371–5375 (1989).
12. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
13. Kojima, H., Muto, E., Higuchi, H. & Yanagida, T. Mechanics of single kinesin molecules measured by optical trapping nanometry. *Biophys. J.* **73**, 2012–2022 (1997).
14. Allersma, M. W., Gittes, F., deCastro, M. J., Stewart, R. J. & Schmidt, C. F. Two-dimensional tracking of *ncd* motility by back focal plane interferometry. *Biophys. J.* **74**, 1074–1085 (1998).
15. Pralle, A., Prummer, M., Florin, E. L., Stelzer, E. H. & Horber, J. K. Three-dimensional high-resolution particle tracking for optical tweezers by forward scattered light. *Microsc. Res. Tech.* **44**, 378–386 (1999).
16. Mason, T. G., Ganesan, K., Van Zanten, J. H., Wirtz, D. & Kuo, S. C. Particle tracking microrheology of complex fluids. *Phys. Rev. Lett.* **79**, 3282–3285 (1997).
17. Yamada, S., Wirtz, D. & Kuo, S. C. Mechanics of living cells measured by laser tracking microrheology. *Biophys. J.* **78**, 1736–1747 (2000).
18. McGrath, J. L., Hartwig, J. H. & Kuo, S. C. The mechanics of F-actin microenvironments depend on the chemistry of probing surfaces. *Biophys. J.* (in the press) (2000).
19. Gerbal, F. et al. On the 'Listeria' propulsion mechanism. *Pramana. J. Phys.* **53**, 155–170 (1999).
20. Kuo, S. C., Gelles, J., Steuer, E. & Sheetz, M. P. A model for kinesin movement from nanometer-level movements of kinesin and cytoplasmic dynein and force measurements. *J. Cell Sci.* **14** (suppl.), 135–138 (1991).
21. Olbris, D. J. & Herzfeld, J. Reconstitution of *Listeria* motility: implications for the mechanism of force transduction. *Biochim. Biophys. Acta* **1495**, 140–149 (2000).
22. Gerbal, F. et al. Measurement of the elasticity of the actin tail of *Listeria monocytogenes*. *Eur. Biophys. J.* **29**, 134–140 (2000).
23. Reif, F. *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).
24. Qian, H., Sheetz, M. P. & Elson, E. L. Single particle tracking: Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* **60**, 910–921 (1991).
25. Molloy, J. E., Burns, J. E., Kendrick-Jones, J., Tregear, R. T. & White, D. C. Movement and force produced by a single myosin head. *Nature* **378**, 209–212 (1995).
26. Mehta, A. D., Finer, J. T. & Spudis, J. A. Detection of single molecule interactions using correlated thermal diffusion. *Proc. Natl Acad. Sci. USA* **94**, 7927–7931 (1997).
27. Dupuis, D. E., Guilford, W. H., Wu, J. & Warshaw, D. M. Actin filament mechanics in the laser trap. *J. Muscle Res. Cell Motil.* **18**, 17–30 (1997).
28. Portnoy, D. A., Jacks, P. S. & Hinrichs, D. J. Role of hemolysin for the intracellular growth of *Listeria monocytogenes*. *J. Exp. Med.* **167**, 1459–1471 (1988).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank L. Machesky for J774 cells, and D. Portnoy and J. Skoble for *Listeria* strains and general advice. We also thank the National Science Foundation, the National Institutes of Health and the Whitaker Foundation for their support. J.L.M. was supported as a BME Distinguished Postdoctoral Fellow, partially funded by the Whitaker Foundation.

Correspondence and requests for materials should be addressed to S.C.K. (e-mail: skuo@bme.jhu.edu).

Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin

Luca Pellegrini*, David F. Burke*, Frank von Delft*, Barbara Mulloy† & Tom L. Blundell*

* Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, UK

† National Institute for Biological Standards and Control, South Mimms, Potters Bar, EN6 3QG, UK

Fibroblast growth factors (FGFs) are a large family of structurally related proteins with a wide range of physiological and pathological activities¹. Signal transduction requires association of FGF with its receptor tyrosine kinase (FGFR)² and heparan sulphate proteoglycan in a specific complex on the cell surface. Direct involvement of the heparan sulphate glycosaminoglycan polysaccharide in the molecular association between FGF and its receptor is essential for biological activity^{3–5}. Although crystal structures of

binary complexes of FGF–heparin^{6,7} and FGF–FGFR^{8,9} have been described, the molecular architecture of the FGF signalling complex has not been elucidated. Here we report the crystal structure of the FGFR2 ectodomain in a dimeric form that is induced by simultaneous binding to FGF1 and a heparin decasaccharide. The complex is assembled around a central heparin molecule linking two FGF1 ligands into a dimer that bridges between two receptor chains. The asymmetric heparin binding involves contacts with both FGF1 molecules but only one receptor chain. The structure of the FGF1–FGFR2–heparin ternary complex provides a structural basis for the essential role of heparan sulphate in FGF signalling.

We reconstituted the ligand-binding extracellular region of human FGFR2 (splice variant IIIc) as a ternary complex with human FGF1 and a heparin decasaccharide by mixing the three components in a 2:2:1 stoichiometric ratio. The ternary complex eluted as a single peak at the expected relative molecular mass of 83,600 (*M*_r 83.6K) on size-exclusion chromatography. The complex was crystallized and its structure determined by X-ray crystallography at 2.8 Å using multiple anomalous dispersion based on selenomethionine substitution of the receptor moiety (Table 1).

Analysis of the crystals shows that one heparin decasaccharide links two FGF1 molecules, each of which binds a receptor ectodomain, unambiguously defining a heteropentameric assembly, consistent with the *in vitro* reconstitution (Fig. 1a, b). The complex can alternatively be considered as two 1:1 FGF1–FGFR2 complexes associated through interaction with the heparin. The ligand-binding fragment of FGFR2 comprises two immunoglobulin-like domains, D2 (residues 152–249) and D3 (residues 254–360), separated by a short linker (residues 250–253)^{8,9} (Fig. 1c). D2 belongs to the I class, whereas D3 approximates to a class C2 fold. FGF1 adopts the well-characterized β-trefoil fold¹⁰. The two receptor chains in the ternary complex show marked bends in the linker regions, with angles of 106° and 87° between D2 and D3. The two protein halves of the complex are related by an approximate two-fold symmetry axis, which we expect would lie perpendicular to the membrane in the cell. The root-mean square deviation (r.m.s.d.) for the Cα superposition of the two halves of the complex is 2.8 Å, whereas the r.m.s.d. for the 230 Cα atoms of FGF1 and D2 is 1.0 Å.

There are two 1:1 FGF–FGFR binary complexes within the heteropentamer, which we refer to as A and B. In A both ligand and receptor interact with heparin, whereas in B only the ligand is in contact with heparin. The entire length of the protein-bound heparin decasaccharide is visible in the experimental electron-density map (Fig. 2a). Heparin lies between and interacts with the two FGF1 molecules, but is tilted away from the quasi-dyad so that it extends past the ligand dimer to interact only with receptor ectodomain A (Figs 1a, b and 3a). The heparin helix presents a marked distortion that has not been observed in previous structural studies of FGF–heparin complexes^{6,7}. The helical axis shows a 34° kink between disaccharides 2 and 3, accompanied by a decreased helical rise between disaccharide 2 and 3 (7.1 Å) and between 3 and 4 (7.9 Å) (for numbering of heparin disaccharide units see Fig. 2b). The total surface area buried at the protein–heparin interface is large at 2,240 Å² (617 Å² with FGF1 and 631 Å² with domain D2 in A; 992 Å² with FGF1 in B). van der Waals contacts contribute substantially to the protein–heparin interactions as suggested previously¹¹. FGF1 binding to heparin is mediated by a set of basic and polar residues, Asn 18, Lys 112, Lys 113, Asn 114, Lys 118, Arg 119, Arg 122, Gln 127 and Lys 128, which were identified in previous structural and mutagenesis experiments^{6,7,11}. FGF1 in binary complex A interacts with six monosaccharides, from iduronic acid in disaccharide 1 (IdoA-1) to glucosamine in disaccharide 4 (GlcN-4), whereas FGF1 in B interacts with five monosaccharides, from GlcN-1 to GlcN-3 (Fig. 2b).

In the structure of the ternary complex, heparin links two FGF1 molecules into a dimer that lacks a protein–protein interface, in a

similar fashion to that previously reported⁷, but with significant differences in the relative orientation of FGF and heparin. Superposition of PDB entry 2AXM (ref. 7) with the heparin-linked FGF1 dimer of the heteropentamer shows that a good agreement is obtained for FGF1 in A and the heparin. However, FGF1 in B is rotated by about 120° relative to the second FGF1 protomer of 2AXM around an axis roughly aligned with the pseudo three-fold axis of the β-trefoil (Fig. 3b). FGF1 in B is thus able to make an

additional set of contacts with heparin, while remaining in contact with it through the heparin-binding loop. Conserved Trp 107 is central to this new heparin-binding site (Fig. 2c). An FGF2 mutant, in which the tryptophan residue was mutated to alanine, showed unaltered affinity for the receptor, but a marked reduction in mitogenic potency¹². It was originally proposed that the amino-acid sequence of the loop containing the tryptophan residue constituted a secondary receptor-binding site¹². Our structure

Table 1 Crystallographic analysis

Diffraction data (space group, $P6_2$; $a = b = 195.90 \text{ \AA}$, $c = 68.86 \text{ \AA}$)								
Data set	Resolution (\AA)	Wavelength (\AA)	Reflections (unique)	Redundancy	Completeness (outer shell)	R_{sym} (%)† (outer shell)	$I/\sigma(I)$ (outer shell)	Beamline
Se–Met (remote)	3.5	0.93928	281,374 (37,018*)	7.6	99.8 (100.0)	11.2 (36.8)	17.1 (3.5)	19-ID
Se–Met (peak)	3.5	0.97924	281,407 (37,767*	7.7	99.9 (100.0)	11.8 (43.0)	14.8 (2.6)	19-ID
Se–Met (edge)	3.5	0.97940	282,931 (36,963*)	7.5	99.9 (100.0)	10.8 (26.1)	13.3 (5.6)	19-ID
Native	2.8	1.00	127,885 (36,432)	3.5	98.3 (100.0)	8.4 (47.4)	13.7 (1.7)	5.2R
MAD phasing	Se–Met (remote)	Se–Met (peak)	Se–Met (edge)					
Rcullis‡ (iso/ano)	–0.90	0.61/0.74	0.70/0.77					
Phasing power§ (iso/ano)	–0.78	1.79/1.86	1.45/1.85					
Refinement								
Resolution (\AA)	Reflections	Number of non-H atoms	R factor (%)¶	Free R factor (%)	Average B (\AA ²)	R.m.s.d. bonds (\AA)	R.m.s.d. angles (degrees)	R.m.s.d. bonded B (\AA ²)
24.2–2.8	36,348	5,247	24.2	28.0	69.1	0.009	1.52	5.36

* For MAD data, the Bijvoet pairs were not merged.

† $R_{\text{sym}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$

‡ R_{cullis} as defined in SHARP.

§ Phasing power as defined in SHARP.

|| Statistics for data with $F > 3\sigma$.

|| R factor = $\sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$

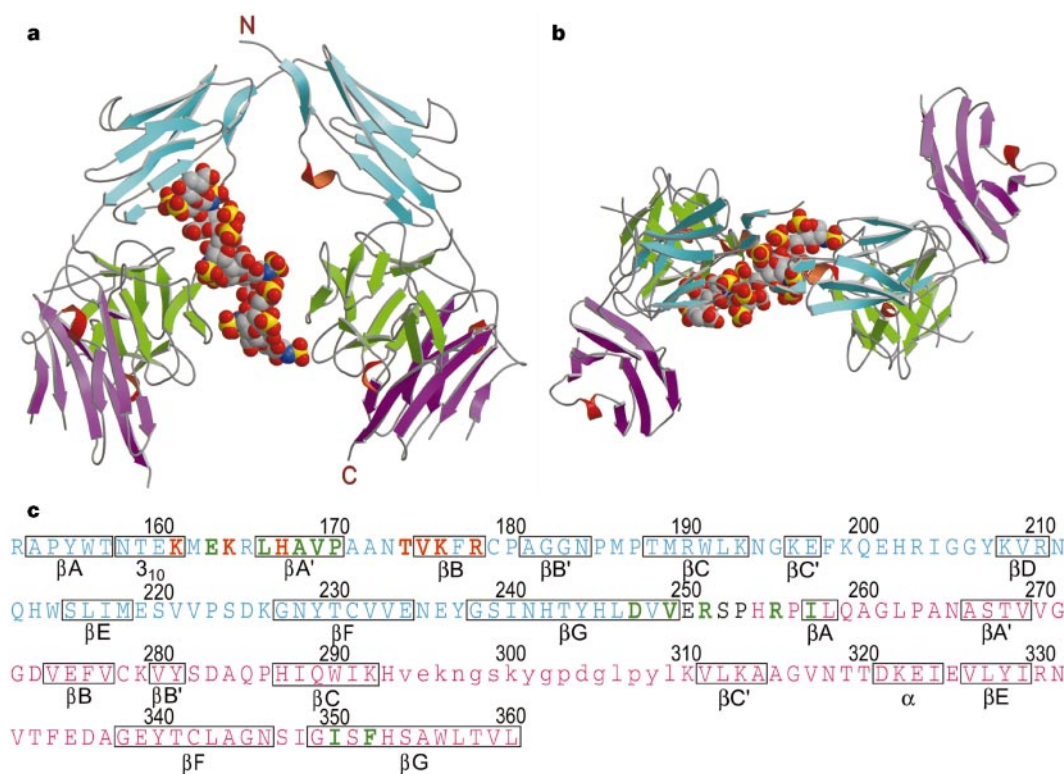


Figure 1 The FGF1–FGFR2–heparin complex. **a**, View perpendicular to the approximate dyad of the complex. FGFR2 domains 2 (D2) and 3 (D3) are cyan and magenta, respectively, and FGF1 is green. The heparin molecule is in CPK representation. **b**, View along the dyad. **c**, Amino-acid sequence of the ligand-binding region of human FGFR2.

Domains D2 and D3 are coloured as in **a**. Residues that interact with heparin and FGF1 are highlighted in red and green, respectively. Regions of secondary structure are boxed. β-strands, α-helices and 3₁₀ helices are indicated below the relative sequence. D3 residues 294–309, which were disordered in the crystals, are in lower case.

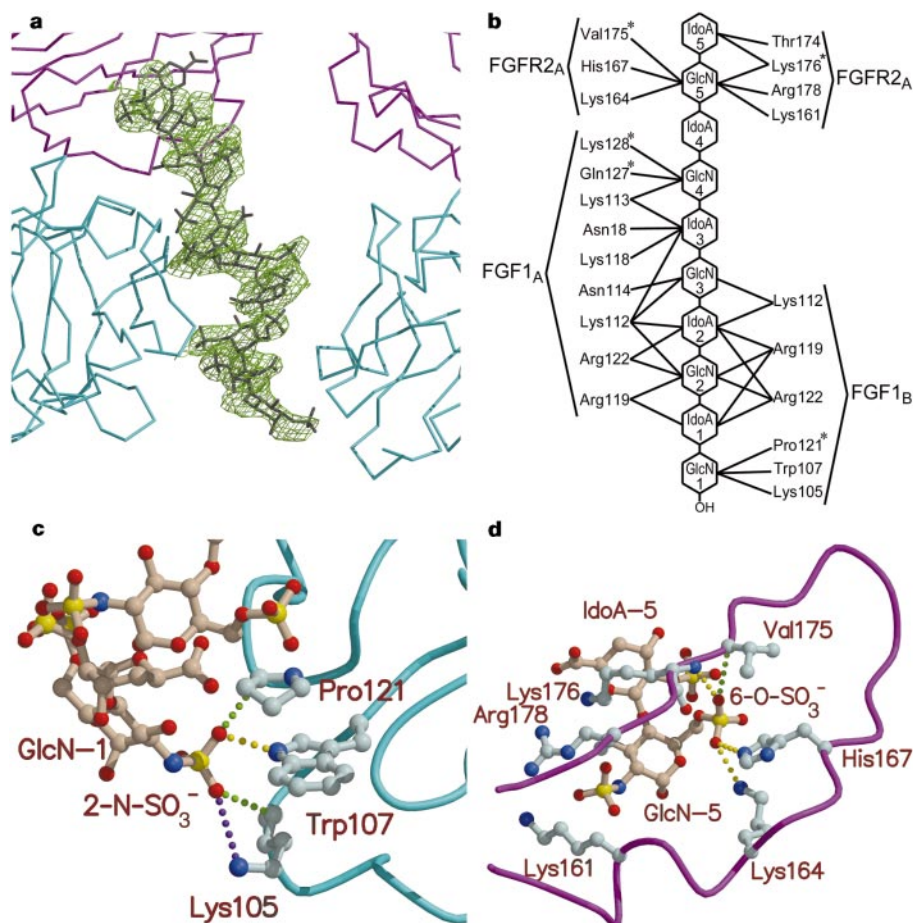


Figure 2 Heparin-protein interactions in the FGF1-FGFR2-heparin complex.

a, Experimental 3.0 Å electron density of the heparin deca-saccharide contoured at 1.2 σ . **b**, Summary of heparin-protein interactions. The five disaccharides of the heparin deca-saccharide are numbered from the reducing end, indicated by a hydroxyl group. Each disaccharide begins with glucosamine (GlcN) followed by iduronic acid (IdoA). Solid lines represent interactions between amino acids and monosaccharide(s). Heparin comes in

close contact with the backbone atoms of amino acids marked by an asterisk. **c**, Details of the FGF1-heparin interaction in binary complex B. Heparin and amino-acid side chains are shown as a ball-and-stick model and the protein backbone as a coil. Dashed lines represent hydrogen bonds (yellow), van der Waals contacts (green) and electrostatic interactions (purple). **d**, FGF2-heparin interactions. Only heparin disaccharide 5 is depicted for clarity. Dashed lines coloured as in **c**.

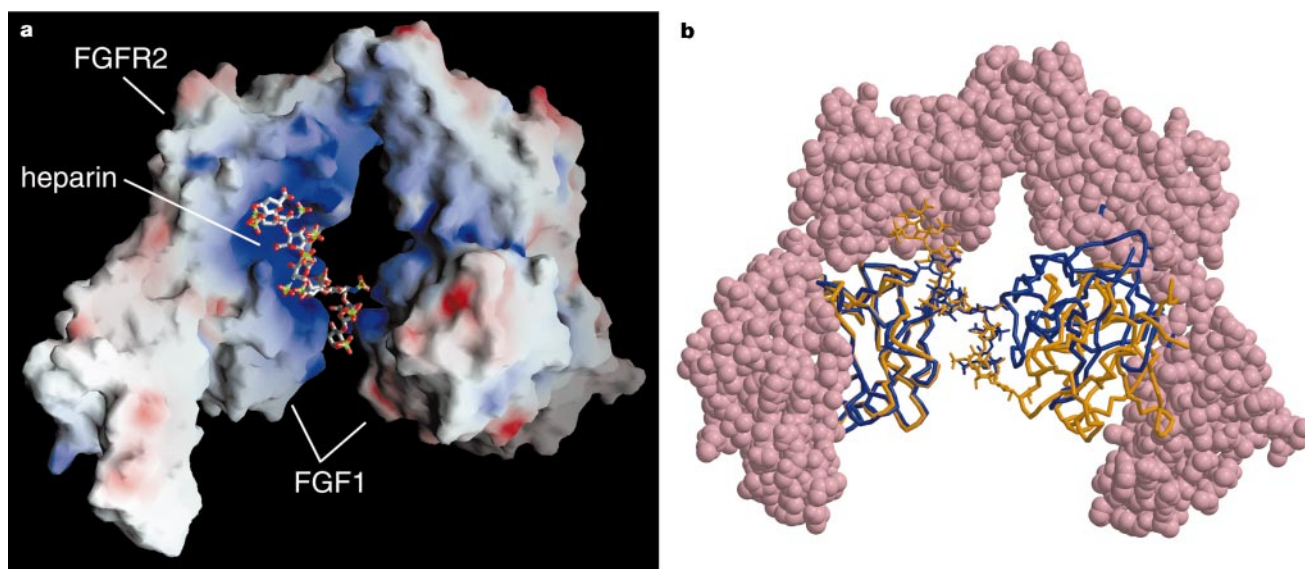


Figure 3 Overall architecture of the FGF1-FGFR2-heparin complex. **a**, Electrostatic potential of the heparin-binding site in the FGF1-FGFR2-heparin complex, mapped onto a molecular surface rendition of the complex. The heparin is shown as a stick model. In the ternary complex, the heparin deca-saccharide is surrounded by regions of positive

charge (blue represents +20e per Å²). **b**, Superposition of heparin-linked FGF1 dimers. The dimer from the FGF1-FGFR2-heparin complex is yellow, the PDB entry 2AXM is blue. The FGFR2 ectodomains are shown in CPK representation.

shows that the N-sulphate group of GlcN-1 is hydrogen bonded to Trp 107 and comes in close contact with Pro 121 and Lys 105, which are stacked above and below the tryptophan side chain (Fig. 2c). This interaction might be important for the correct positioning of

FGF1 relative to heparin and thus for the recruitment of a second binary complex into an active heteropentamer.

The heparin-binding region of FGFR2 observed in the structure of the ternary complex comprises residues in the highly conserved

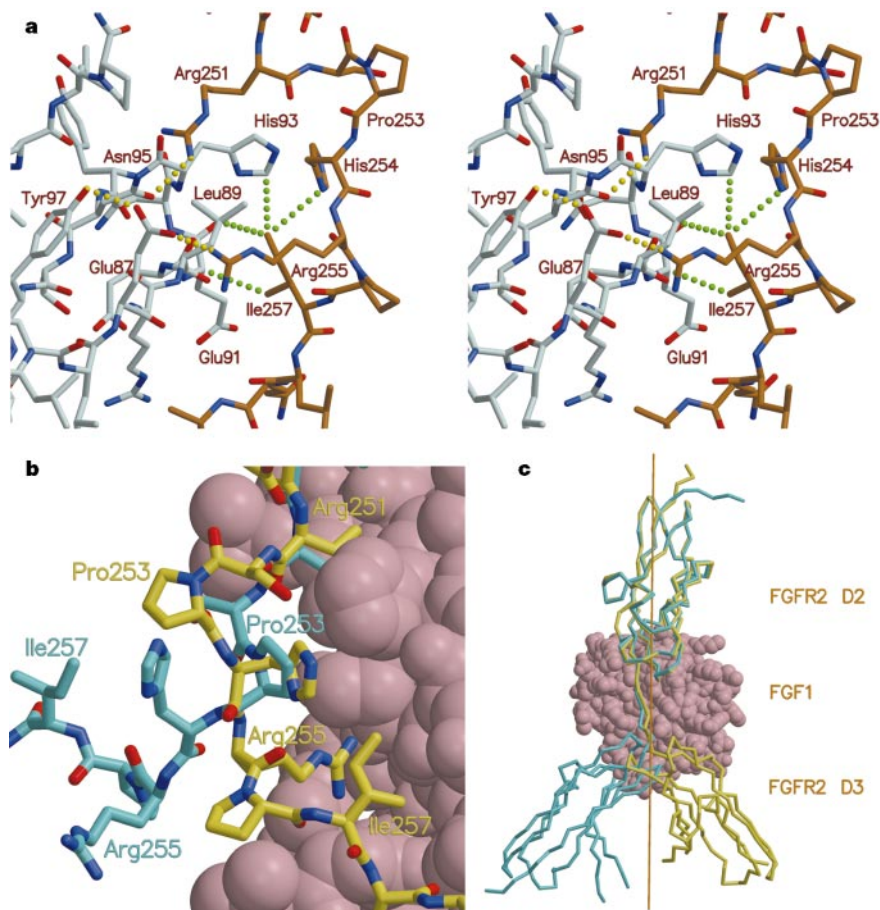


Figure 4 Ligand-receptor interactions in FGF-FGFR complexes with and without heparin. **a**, Stereoview of the ligand-receptor interface in the FGF1-FGFR2-heparin complex. The receptor is brown and the ligand is light grey. Dashed lines represent hydrogen bonds (yellow) and van der Waals contacts (green). **b**, Close-up of FGFR2 sequence R251-S-P-H-R-P-I257, illustrating the different conformation of the polypeptide chain in FGF-FGFR complexes with (binary complex A; yellow) and without

(PDB entry 1DJS; cyan) heparin. FGF1 is in CPK representation. **c**, Superposition of FGFR2 ectodomains showing the different orientation of D3 in FGF-FGFR complexes with and without heparin. The Cα trace of FGFR2 ectodomain from binary complex A (yellow) is superimposed on that of PDB entry 1DJS (cyan). FGF1 is also shown in CPK representation. The rotation axis that relates the two D3 domains is drawn.

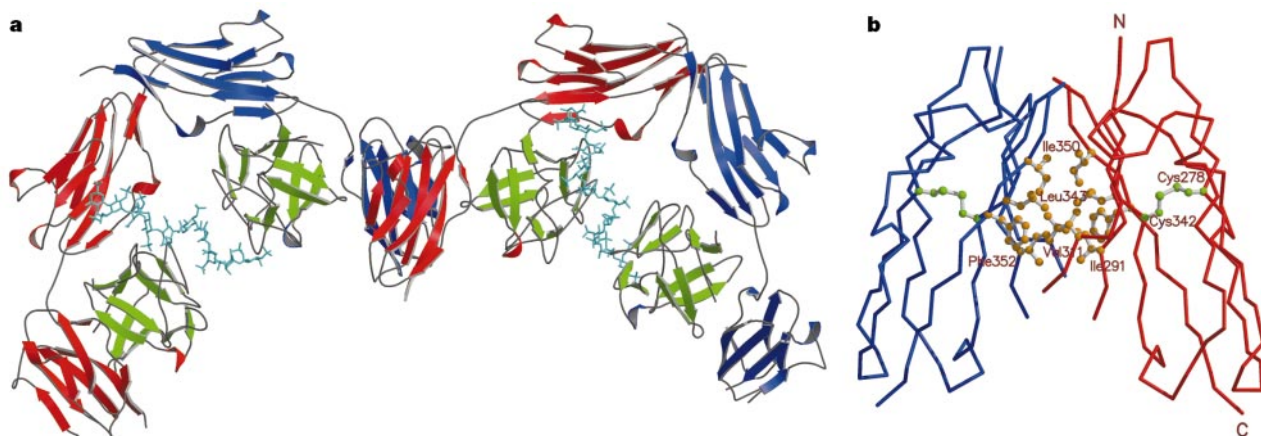


Figure 5 Higher order FGFR oligomerization as observed in crystals of the FGF1-FGFR2-heparin complex. **a**, Crystal packing of neighbouring heteropentamers. Linking successive heteropentamers together by apposition of D3 domains generates the hexagonal symmetry in the crystals. FGFR2 ectodomains are red and blue, FGF1 is green

and heparin is cyan. **b**, Face-to-face interaction between FGFR D3 domains belonging to neighbouring heteropentamers. The protein backbone is depicted as a Cα trace, the hydrophobic residues at the D3-D3 interface and the disulphide bridge of each D3 domain are shown as ball-and-stick models.

sequence K161-M-E-K-R-L-H-A-V-P-A-A-N-T-V-K-F-R178, which has been identified as being essential for heparin binding¹³ (Figs 1c and 2b, d). All basic residues in this sequence, with the exception of Arg 165, are involved in interactions with the sulphate groups of GlcN-5 or the carboxylate group of IdoA-5. Both sugars in disaccharide 5 come into close contact with the receptor surface. The 6-O-sulphate group of GlcN-5 is involved in a particularly large range of interactions, including hydrogen bonds to Lys 164, His 167 and the backbone amide of Lys 176, and a van der Waals contact with Thr 174 and Val 175 (Fig. 2d). This is in agreement with reports that 6-O-desulphated heparin retains its ability to bind FGF, but blocks receptor binding and mitogenic activity¹⁴. The role of Lys 164 in neutralizing the 6-O-sulphate group of GlcN-5 appears particularly important as the sulphate group is partially buried.

The structure of the ternary complex shows that a heparin span of eight hexoses would still be capable of crosslinking ligand and receptor, through interactions of Lys 161 with the 2-O-sulphate group of IdoA-4 and Lys 164 with the carboxylate function of IdoA-4, albeit presumably with reduced efficacy. Our structure predicts that maximization of the heparin–protein interactions in the ternary complex would be achieved with a heparin fragment spanning twelve sugars: a dodecasaccharide should be able to make an additional set of contacts with Lys 208 and Arg 210 on β -strand D of domain D2 (Fig. 1c). These observations are in agreement with biochemical data that indicate that an octasaccharide is the minimal heparin length required to support FGF physiological activity¹⁵ and that a dodecasaccharide is equivalent to whole heparin in FGF1 proliferation potency¹⁶.

Protein–protein interactions in the ternary complex are dominated by ligand binding of each receptor chain. FGF1 binding to the receptor involves contacts with D2, the linker region and residues 254–260 of D3, which create a continuous interface burying a total of 1,700 Å² surface area. This is consistent with biochemical evidence that narrows the minimal ligand-binding region of FGFR to a fragment comprising D2, the interdomain sequence and an amino-terminal sequence of D3 (ref. 17). Contacts between FGF1 and domain D2 are predominantly hydrophobic and involve residues Tyr 15, Gly 20, Phe 22, Tyr 94, Leu 133 and Leu 135 of the ligand, and residues Lys 164, Leu 166, Ala 168, Val 169 and Pro 170 of D2. Interactions of an electrostatic nature are those of Glu 163 and Asp 247 of FGFR2 with Arg 35 and Arg 37 of FGF1, respectively. These observations closely resemble those made in previous biochemical and structural studies^{8,9,12}.

In the linker between D2 and D3, the guanidinium moiety of invariant Arg 251 is surrounded by three hydrophobic FGF1 residues, Leu 89, Leu 133 and Pro 134, and is hydrogen bonded to the main chain of FGF1 His 93. These interactions immobilize the arginine side chain in the correct position to form a crucial hydrogen bond with the conserved Asn 95 of FGF1 (Fig. 4a). The importance of this interaction is highlighted by mutagenesis data that show a 400-fold reduction in receptor affinity when the equivalent asparagine in FGF2 is replaced by alanine¹⁸.

In domain D3, the conserved Arg 255 forms an ion pair with the invariant Glu 87 of FGF1 (Fig. 4a). Involvement of Glu 87 in a charge–charge interaction has been demonstrated by mutagenesis experiments that show a 10-fold reduction in receptor-binding affinity when glutamate is replaced by glutamine in FGF2 (ref. 19). Glu 87 is favourably positioned for interaction with Arg 255 by a hydrogen bond with neighbouring Tyr 97. Invariant Ile 257 of D3 packs hydrophobically against Leu 89 and His 93 and the main-chain atoms linking Glu 90 and Glu 91 of FGF1 and His 254 of the receptor (Fig. 4a). The only spliceform-specific FGF1–D3 interaction in the ternary complex is a hydrophobic contact between Val 51 of FGF1 and Ile 350 and Phe 352 of the receptor. The presence of a single contact with the spliceform-specific region of FGFR explains the lack of specificity of FGF1 towards different receptor isoforms.

Comparison of FGF–FGFR crystal structures with and without

heparin, obtained under similar crystallization conditions, shows that protein–protein interactions involved in crystal packing are partly conserved. The 3₁ screw axis in our crystals brings two D2 domains from different complexes together, forming a cleft between them in a fashion reminiscent of the canyon-forming interaction observed in crystals of FGF–FGFR complexes without heparin^{8,9}. This cleft was proposed to be the heparin-binding site in the FGF–FGFR–heparin ternary complex^{8,9}. In our structure a disaccharide of the heparin occupies the entry to the cleft, leaving it otherwise empty. Furthermore, analysis of crystal packing in FGF–FGFR complexes without heparin shows that the canyon-forming interaction between FGFR ectodomains is not observed when crystallization is achieved in a low salt buffer (compare PDB entry 1CVS with entries 1EVT and 1EV2). The crystallization process, and not heparin, is therefore likely to be responsible for the association of two FGFR molecules in a dimer with a canyon-like cleft.

Relative to its position in FGF–FGFR complexes without heparin^{8,9}, receptor domain D3 is swivelled around the linker region by 174.5° in A and by 163.3° in B (Fig. 4c). In the crystals neighbouring heteropentamers related by the 6₂ screw axis are linked through face-to-face packing of two D3 domains, involving a cluster of conserved hydrophobic residues, Ile 291, Val 311, Leu 343, Ile 350 and Phe 352 on β -strands G, F, C, C' (Fig. 5). It is possible that such an interaction may be involved in higher order FGFR oligomerization and the clustering observed in the focal adhesion complex²⁰.

The different receptor conformations observed in FGF–FGFR complexes with and without heparin suggest a mechanism to regulate receptor function. FGF and FGFR may initially associate in a binary complex with the receptor ectodomain in the conformation observed previously^{8,9}. Heparin binding would cause the D3 domain to rotate into the position observed in our structure through conformational changes in the linker, possibly involving Pro 253, which is in *trans* conformation in FGF–FGFR complexes without heparin^{8,9}, but adopts a *cis* conformation in the ternary complex (Fig. 4b). Indirect evidence in support of such a mechanism comes from the observation that the rare Ser252Phe FGFR2 mutation found in Apert syndrome is responsible for a severe phenotype similar to that of recurrent Ser252Trp, whereas Ser252Leu causes a normal phenotype²¹. Aromatic amino acids preceding a proline induce a high fraction of *cis* isomer in test peptides²². The severity of the Apert phenotype might thus be determined by the propensity of the Ser-252 mutation to stabilize the *cis* isomeric form of Pro 253.

We propose that our structure represents the active dimeric state of FGFR when bound to FGF and heparin. The role of heparin in receptor dimerization appears to be critical, as little protein–protein interface is present between the two halves of the heteropentamer. The absence of significant contacts between the two halves of the heteropentamer would facilitate heterodimerization of different ligand and receptor types. The asymmetric binding of heparin observed in the crystal is consistent with a stepwise assembly of the FGF signalling complex. The structure also suggests mechanisms for receptor activation, based on a heparin-induced conformational change in the receptor, and for higher order receptor oligomerization. The universality of the architecture of the FGF–FGFR–heparin signalling complex that emerges from this study will now have to be confirmed by structural determination of ternary complexes involving different members of the FGF family of ligands and receptors. □

Methods

Preparation of recombinant protein

We subcloned full-length human FGF1 complementary DNA into expression vector pET14a (Novagen), overexpressed the protein in a soluble form in *Escherichia coli* BL21(DE3) strain and purified it to homogeneity over a heparin Sepharose affinity column (Pharmacia). We subcloned human FGFR2 cDNA (splice variant IIIc) comprising

the extracellular immunoglobulin-like domains D2 and D3 (residues 148–366) into expression vector pET3a (Novagen), overexpressed it in *E. coli* BL21(DE3) strain as inclusion bodies, and refolded and purified it to homogeneity by a combination of standard chromatography techniques.

Preparation of heparin decasaccharide

Monodisperse heparin decasaccharide was prepared from a sample of heparin remaining from the 2nd International Standard. This material is highly homogeneous, containing more than 90% of the predominant repeating unit of heparin, the disaccharide α -L-IdoA(2OSO₃⁻)-(1→4)- α -D-GlcNSO₃⁻(6OSO₃⁻)-(1-. The parent heparin was partially degraded by brief incubation with heparin lyase 1 as described²³. The resulting oligosaccharides were separated by size-exclusion chromatography on a Biogel P10 column (1.6 × 100 cm) (BioRad) using 2% ammonium bicarbonate as eluent at 18 ml h⁻¹. Fractions (6 ml) were monitored by absorbance at 235 nm and pooled appropriately. The size of each oligosaccharide was confirmed by analytical size-exclusion chromatography.

Preparation of ternary complex and crystallization

The ternary complex of FGF1, FGFR2 and heparin was reconstituted by mixing the three components in 2:2:1 stoichiometric ratios and isolated by size-exclusion chromatography over a Superdex 200 HR 10/30 column (Pharmacia) as a single peak of *M*_r 83.6K. Crystals were grown at 18° by the hanging drop method, mixing a 0.1 mM solution of ternary complex in 150 mM NaCl, 10 mM HEPES pH 7.2, 25 mM MgCl₂ with equal volumes of well solution containing 1.0 M Li₂SO₄, 0.1 M Tris-HCl pH 8.5, 10 mM NiSO₄. The crystals belong to space group P6₂ with unit-cell dimensions *a* = *b* = 195.90 Å, *c* = 68.86 Å, and contain one 2:2:1 complex per asymmetric unit with 72% solvent content. Crystals grown in the absence of NiSO₄ are isomorphous with the ones produced under the original conditions, but were too small for high-resolution X-ray analysis.

Phasing and refinement

The structure was determined by the multiple anomalous dispersion (MAD) method using selenomethionyl FGFR2 protein. MAD data at three different wavelengths near the Se K edge (Table 1) were collected at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source, Chicago, US. The images were integrated and the intensities merged with HKL2000 (ref. 24). The positions of the eight selenium atoms in the asymmetric unit were determined with SHELX-97 (ref. 25). Phases were calculated in SHARP²⁶ and improved by SOLOMON²⁷. The resulting electron-density map at 3.5 Å resolution clearly showed the location of all components of the ternary complex. A homology-based model for FGFR2 domain D2 and a crystallographic FGF1 model (PDB entry 2AFG) were initially positioned in the experimental map. After a few cycles of phase combination and extension, a heparin decasaccharide²⁸ and a homology-based model for FGFR2 domain D3 were added. The structure was refined against native data measured at beamline 5.2 R of Elettra, Trieste, Italy, to 2.8 Å resolution. Refinement of the crystallographic model was carried out in CNS²⁹, alternated with manual rebuilding in O³⁰. Tight non-crystallographic symmetry restraints were individually applied to FGF1 and the D2 and D3 receptor domains throughout the refinement.

NMR studies of free heparin in solution show that the two main conformations of the iduronate ring, ¹C₄ chair and ²S₀ skew boat, are in equilibrium²⁸, and crystallographic studies of FGF2–heparin complexes reveal that both forms are present in protein-bound heparin²⁸. The pyranose rings of the iduronate moieties in the heparin were kept in the ²S₀ skew boat conformation until a late phase of the refinement, when inspection of the density map warranted a switch to the ¹C₄ chair conformation for IdoA-1. Our assignment of the iduronate ring conformations must be considered tentative owing to the limited resolution of our study. Heparin lyase I digestion leaves an unsaturated iduronic acid at the non-reducing end of the heparin oligosaccharide. This chemical modification was not introduced in our model of the heparin molecule.

The refined model comprises 5,001 protein atoms, 175 heparin atoms, 41 water molecules, 5 sulphate ions and 5 nickel ions. 1.1% of the protein residues are in the generously allowed region and none in the disallowed region of the Ramachandran plot. Residues 1–9 (1–8 in B) and 139–140 of FGF1 and residues 148, 294–308 (294–309 in B) and 361–366 of FGFR2 are not visible in the electron-density map and have not been included in the final model. The quality of the electron-density map for residues 330–338 and 356–360 of FGFR2 is poor, indicating that they are partially disordered in the crystal. Residues for which no side-chain density was present were modelled as alanine during refinement.

Surface area accessibility calculations were carried out as implemented in CNS²⁹. Molecular superpositions and r.m.s.d. calculations were performed with LSQMAN. Figures were prepared with Molscript, Raster3D and GRASP.

Received 10 April; accepted 15 September 2000.

- Galzie, Z., Kinsella, A. R. & Smith, J. A. Fibroblast growth factors and their receptors. *Biochem. Cell Biol.* **75**, 669–685 (1997).
- Johnson, D. E. & Williams, L. T. Structural and functional diversity in the FGF receptor multigene family. *Adv. Cancer Res.* **60**, 1–41 (1993).
- Rapraeger, A. C., Krufka, A. & Olwin, B. B. Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science* **252**, 1705–1708 (1991).

- Spivak-Kroizman, T. *et al.* Heparin-induced oligomerization of FGF molecules is responsible for FGF receptor dimerization, activation and cell proliferation. *Cell* **79**, 1015–1024 (1994).
- Lin, X., Buff, E. M., Perrimon, N. & Michelson, A. M. Heparan sulfate proteoglycans are essential for FGF receptor signaling during *Drosophila* embryonic development. *Development* **126**, 3715–3723 (1999).
- Faham, S., Hileman, R. E., Fromm, J. R., Linhardt, R. J. & Rees, D. C. Heparin structure and interactions with basic fibroblast growth factor. *Science* **271**, 1116–1120 (1996).
- DiGabriele, A. *et al.* Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature* **393**, 812–817 (1998).
- Plotnikov, A. N., Schlessinger, J., Hubbard, S. R. & Mohammadi, M. Structural basis for FGF receptor dimerization and activation. *Cell* **98**, 641–650 (1999).
- Stauber, D., DiGabriele, A. & Hendrickson, W. Structural Interactions of fibroblast growth factor receptor with its ligands. *Proc. Natl Acad. Sci. USA* **97**, 49–54 (2000).
- Zhu, X. *et al.* Three-dimensional structure of acidic and basic fibroblast growth factor. *Science* **251**, 90–93 (1991).
- Thompson, L. D., Pantoliano, M. W. & Springer, B. A. Energetic characterization of the basic fibroblast growth factor-heparin interaction: identification of the heparin binding domain. *Biochemistry* **33**, 3831–3840 (1994).
- Springer, B. A. *et al.* Identification and concerted function of two receptor binding surfaces on basic fibroblast growth factor required for mitogenesis. *J. Biol. Chem.* **269**, 26879–26884 (1994).
- Kan, M. *et al.* An essential heparin-binding domain in the fibroblast growth factor receptor kinase. *Science* **259**, 1918–1921 (1993).
- Guimond, S., Maccarana, M., Olwin, B. B., Lindahl, U. & Rapraeger, A. C. Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4. *J. Biol. Chem.* **268**, 23906–23914 (1993).
- Ornitz, D. M. *et al.* Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells. *Mol. Cell Biol.* **12**, 240–247 (1992).
- Sudhalter, J., Folkman, J., Svahn, C. M., Bergendal, K. & D'Amore, P. A. Importance of size, sulfation, and anticoagulant activity in the potentiation of acidic fibroblast growth factor by heparin. *J. Biol. Chem.* **264**, 6892–6897 (1989).
- Wang, F., Kan, M., Xu, J., Yan, G. & McKeehan, W. L. Ligand-specific structural domains in the fibroblast growth factor receptor. *J. Biol. Chem.* **270**, 10222–10230 (1995).
- Zhu, H. *et al.* Analysis of high-affinity binding determinants in the receptor binding epitope of basic fibroblast growth factor. *Protein Eng.* **10**, 417–421 (1997).
- Zhu, H. *et al.* Glu-96 of basic fibroblast growth factor is essential for high affinity receptor binding. Identification by structure-based site-directed mutagenesis. *J. Biol. Chem.* **270**, 21869–21874 (1995).
- Plopper, G. E., McNamee, H. P., Dike, L. E., Bojanowski, K. & Ingber, D. E. Convergence of integrin and growth factor receptor signaling pathways within the focal adhesion complex. *Mol. Biol. Cell* **6**, 1349–1365 (1995).
- Oldridge, M. *et al.* Genotype-phenotype correlation for nucleotide substitutions in the IgII–IgIII linker of FGFR2. *Hum. Mol. Genet.* **6**, 137–143 (1997).
- Reimer, U. *et al.* Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.* **279**, 449–460 (1998).
- Rice, K. G., Kim, Y. S., Grant, A. C., Merchant, Z. M. & Linhardt, R. J. High-performance liquid chromatographic separation of heparin-derived oligosaccharides. *Anal. Biochem.* **150**, 325–331 (1985).
- Otwonowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Sheldrick, G. M. & Schneider, T. R. SHELXL: high resolution refinement. *Methods Enzymol.* **277**, 319–343 (1997).
- La Fortelle, E. d. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
- Mulloy, B., Forster, M. J., Jones, C. & Davies, D. B. N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **293**, 849–858 (1993).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Acknowledgements

We would like to thank P. Gadhave for discussions and early work on FGFR overexpression; D. Chirgadze, E. Parisini and J. Patel for help with X-ray data collection; R. Sanishvili for user support at the SBC beamline; M. Vinkovic for help with the use of SHELX; M. Hyonen for advice on protein refolding and help with the figures; and M. Symmons for advice on production of selenomethionyl protein. We are grateful to S. Malcolm for insightful comments on the phenotype of FGFR mutations in Apert syndrome and for the FGFR2 cDNA, and to D. Fernig for the FGF1 cDNA. Heparin lyase I was a gift of K. Johansen, Leo Pharmaceuticals, Denmark. This research was supported by grants from the MRC and the Wellcome Trust. Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science.

Correspondence and requests for materials should be addressed to L.P. (e-mail: luca@cryst.bioc.cam.ac.uk) or T.L.B. (e-mail: tom@cryst.bioc.cam.ac.uk). Coordinates have been deposited in the Protein Data Bank with accession number 1e0o.

New for the neurosciences

Kainic acid is back on the menu — but no longer from the sea.

S100A6 antibody

DAKO Corporation www.dakousa.com

Proteins involved in neurological disorders

The S100A6 protein belongs to the S100 subfamily, one of the largest subfamilies of the EF-hand Ca^{2+} -binding proteins. S100 proteins are individually regulated in normal and tumour cell lines and tissues. Members of this subfamily are of particular interest because of their deregulated expression in neurological and neoplastic disorders. DAKO's rabbit anti-human S100A6 antibody (No. A5115), for research use, can be used on formalin-fixed, paraffin-embedded tissues, cryostat sections, and cell smears using detection systems such as DAKO's LSAB+ or EnVision+. Antibodies S100A1, A2, A4 and S100B are also available.

Reader Service No. 100

Synphilin-1 antibodies

Chemicon International www.chemicon.com

Polyclonal antibodies to a novel protein

Synphilin-1 is a protein involved in Parkinson's disease and dementia, diseases with Lewy bodies that are characterized by α -synuclein-containing neural inclusions. Two antibodies are available from Chemicon: AB5388 made against a 19-mer and AB5450 against a peptide of 9 amino acids. Earlier this year, Chemicon introduced a new guinea-pig polyclonal antibody to the GABA-B R2 receptor, Catalogue No. AB5394. The antibody is raised against rat GABA-B R2 receptor.

Reader Service No. 101

BrainWare

Tucker-Davis Technologies www.tdt.com

Software for electrophysiological recording

BrainWare is a package for stimulus presentation and extracellular electrophysiological recording, providing neural spike acquisition and analysis on a PC. BrainWare is based on TDT's System II instrumentation for auditory stimulus presentation and is fully compatible with Cambridge Research Systems VSG to create and present visual stimuli. BrainWare provides online cluster cutting for spike discrimination and plots neural responses versus stimulus parameters in real-time, giving immediate feedback about the responsiveness of cells.

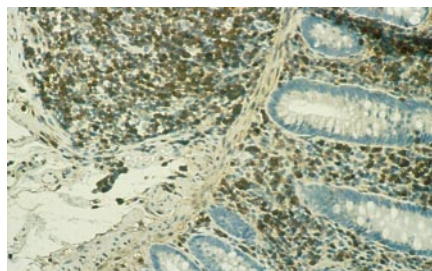
Reader Service No. 102

Antibodies for neurobiology

Zymed www.zymed.com

Family values

Zymed's range of antibodies for the study of neurobiological systems includes antibodies



Sectional interest: S100 in tonsil tissue.

specific for various families of neurotransmitters and their receptors. Over 15 different kinds of purified and concentrated glutamate receptor subunit and splice variant antibodies are available. These purified antibodies are suitable for western blotting, and some selected glutamate receptor subunit-specific antibodies can be used for ELISA, immunoprecipitation and immunohistostaining. Anti-GluR1, 2 and 3, anti-GluR2, anti-GluR2 and 4, anti-GluR3 and anti-glutamate transporter EAAC1 are available as monoclonals. Anti-mGluR is available as polyclonal antibody only.

Reader Service No. 103

M2Bio

Carl Zeiss

www.zeiss.com

A 'new dimension' in GFP imaging

Carl Zeiss, Inc. and Kramer Scientific have launched the M2Bio Fluorescence Stereo microscope, combining stereo macro and compound micro imaging capabilities in a single platform. This system allows the user to sort larger populations of a sample in various stereoscopic modes, $\times 10$ – $\times 105$ magnification, or simply pivot the microscope system into position for high-magnification for more detailed observation at $\times 200$ – $\times 660$. Previously, this task was accomplished by moving a sam-



From macro to micro with the M2Bio.

ple to a different microscope. The addition of fixed shutter operated transmitted light offers hands-free operation, providing a fully functional system for a broad range of applications.

Reader Service No. 104

MACS Cell Isolation Kits

Miltenyi Biotec

www.miltenyibiotec.com

Dendritic cell research

Miltenyi Biotec recommend their latest separation system for study of the three distinct populations of blood dendritic cells (BDCs) that can now be distinguished, based on the expression of BDC antigens. BDCA-2 and BDCA-4 are specifically expressed on plasmacytoid (lymphoid) BDC. BDCA-1 (CD1c) and BDCA-3 discriminate two populations of BDC, which are both, according to the phenotype, of myeloid origin: BDCA-1 (CD1c) is expressed on $\text{CD11c}^{\text{high}}$ $\text{CD123}^{\text{low}}$ BDC and BDCA-3 specifically marks a population that was previously not well described, characterized as $\text{CD11c}^{\text{low}}$ and CD123^- . All three populations can be identified in peripheral blood by fluorescence microscopy or flow cytometry using MACS fluorochrome conjugates.

Reader Service No. 105

ADVERTISEMENTS

**Cross species
Primer design
with
Primer Premier**

✓ Allele specific primers ✓ Degenerate Primers
✓ ClustalW alignment ✓ Pathogenic primers
✓ Restriction enzyme & motif analysis
✓ Nested/multiplex primers and more...

Download free demo from
<http://www.PremierBiosoft.com>

Call 888-847-7494 (Toll free in US)
650-856-2703 (International)

PREMIER
Biosoft International
Intuitive software for the molecular biologist

Look no further

for the best value in

Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 45

READER ENQUIRY NO. 33

new on the market

Levelock LL-2

ALA Scientific Instruments

www.alascience.com

On the level

Fluid-level control for very small chambers is frequently needed in patch-clamp and intra/extracellular recordings (to stabilize bath capacitance), imaging studies (for optical clarity), and in experiments requiring accurate solution flow. With minimal surface disturbance and accuracy to 100 μm , Levelock LL-2 offers fluid-level control for volumes as low as 1 ml and handles flow rates up to 9 ml per min. Electrical noise and mechanical artefacts are minimized for patch-clamp applications.

Reader Service No. 106

Synthetic kainic acid

Sigma-RBI

www.sigma-aldrich.com

As much kainic acid as you can use

Until recently, kainic acid was extracted from the seaweed, *Digena simplex*, which primarily grows off Taiwan. Its main use was as an anti-worming agent, but it is also used as a neurodegenerative agent in animals and as a tool to define the EAAT2 subtype of glutamate transporter. With better antiworming agents available, the main producer of kainic acid withdrew it two years ago, leading to a world shortage. Now Sigma-RBI offer 99% pure synthetic kainic acid that matches the previous specification. The product is glutamate free. Available in 10- and 50-mg sizes, there is no restriction on the quantity you can order.

Reader Service No. 107

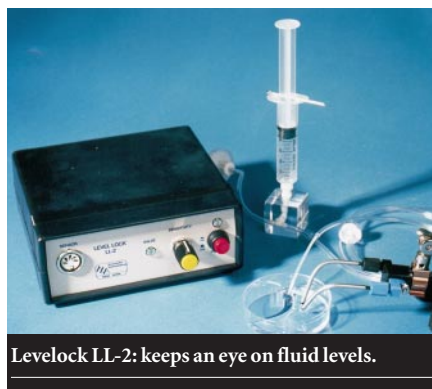
Minicatalogue

Alexis Biochemicals

www.alexis-corp.com

A 20-page mini-catalogue of neurochemicals

This catalogue lists almost 275 reagents for neuropharmacology research. New products and technical notes are featured for major product groups including Ligands for Metabotropic Glutamate Receptors, Vanilloid



Levelock LL-2: keeps an eye on fluid levels.

Receptor Research and Furin Convertase. Each listing contains up-to-date references, including the complete title of the article.

Reader Service No. 108

Cardiac/neuroscience antibodies

Kamiya

www.kamiyabiomedical.com

Anti-ouabain monoclonals

Kamiya Biomedical has introduced monoclonal antibodies to endogenous ouabain-like factor for cardiac and neuroscience research. Antibodies from clones T8B11 and T6D11 recognize the pharmacologically active portion of ouabain and cross-react with ouabain-like K^+ -rival type Na^+ , K^+ -ATPase inhibitory substances consisting of the *cis*-linked D ring and the functional groups at positions C-13, C-14 and C-17. The antibodies neutralize the biological activity of ouabain and endogenous ouabain-like activities *in vivo*. The antibodies are also available conjugated to alkaline phosphatase, biotin, or horseradish peroxidase for immunohistochemical studies.

Reader Service No. 109

CCD camera

Hamamatsu Photonics

hamamatsu.co.uk

A slice of the action

The visualization of individual neurons within thick brain slices is difficult with standard CCD cameras, but Hamamatsu claim the new

C2400-79 CCD video camera with enhanced near-infrared sensitivity is ideal for low-contrast applications such as imaging IR-DIC patch clamping, allowing clear visualization of nerve cells within thick sections. The sensitivity of a standard CCD camera peaks at about 550 nm and extends to 1,100 nm. The C2400-79 has a maximum at 800 nm and extends to 1,300 nm. The C2400-79 combines with Hamamatsu's Argus-20 and Windows software to provide a Patch-ISV imaging function.

Reader Service No. 110

Radiochemicals

American Radiolabeled Chemicals

www.arc-inc.com

Tritium and carbon-14 via new website

The enhanced American Radiolabeled Chemicals website interfaces a catalogue of over 1,600 radiolabelled compounds for life science research, with an online ordering system and full-text search engine. The Highlights section features selected products including fatty acids and ceramides/sphingolipids/phospholipids.

Reader Service No. 111

Confocal microscopy

Bio-Rad

www.microscopy.bio-rad.com

Chondroitin sulphates under the microscope

Bio-Rad's Application Note 29 deals with the use of combined confocal microscopy and specific immunochemistry in work on axonal guidance mechanisms in the vertebrate brain. R. B. Anderson *et al.* studied the effect of chondroitin sulphates on the growth and trajectory of axons which normally grow in the ventral midbrain in *Xenopus* embryos. Findings from intact *Xenopus* brain *in vivo* suggest that chondroitin sulphate proteoglycans regulate axon growth or provide guidance cues, which then modulate the axon's progress and trajectory.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

www.Alphainnotech.com

New Visions in Digital Imaging

Alpha Innotech Corporation

READER ENQUIRY NO. 48

Custom Peptides & Polyclonal Antibodies

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

Research Genetics
U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our homepage at
<http://www.resgen.com>

READER ENQUIRY NO. 36

© 2000 Macmillan Magazines Ltd

BLOCKERS

K⁺ Channel Blocker **E4031**

Arachidonic Acid Cascade Blocker **AA 861**

Wako USA 877-714-1921

Wako

Wako GmbH (49) 02131 3110

www.wako-chem.co.jp

READER ENQUIRY NO. 35